GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 01:02:58 ; Search time 1188 Seconds
(without alignments)
5879.355 Million cell updates/sec

Title: US-09-980-277-1
Perfect score: 240
Sequence: 1 gatctgaatccgaccaaccg.....cgcaattatattttgttctt 240

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44.2	18.4	121	3	SCMRPTAA	M61098 S.mansoni t
2	40.8	17.0	69028	9	AC087493	AC087493 Homo sapi
3	40.8	17.0	150794	9	AC025211	AC025211 Homo sapi
c 4	40.8	17.0	186360	9	AC007982	AC007982 Homo sapi
5	40.8	17.0	218922	2	AC008106	AC008106 Homo sapi
6	40.6	16.9	171744	2	AC095286	AC095286 Rattus no
c 7	39.8	16.6	57075	9	AC080073	AC080073 Homo sapi
c 8	39.2	16.3	180104	9	DJ526N18	AF123462 Homo sapi
c 9	38.4	16.0	166434	2	AC109963	AC109963 Rattus no
c 10	38.4	16.0	192153	2	AC095187	AC095187 Rattus no
c 11	38	15.8	32940	3	AF022972	AF022972 Caenorhab
c 12	37.6	15.7	114473	2	AC111374	AC111374 Rattus no
c 13	37.6	15.7	177720	9	AC007514	AC007514 Homo sapi
c 14	37.6	15.7	190481	2	AC125626	AC125626 Rattus no
15	37.4	15.6	220161	2	AC094816	AC094816 Rattus no
c 16	37.4	15.6	232004	2	AC023584	AC023584 Homo sapi
17	37.2	15.5	68133	2	AC100012	AC100012 Mus muscu
c 18	37.2	15.5	165733	2	AC099146	AC099146 Rattus no
19	37	15.4	168142	2	AC102140	AC102140 Mus muscu
c 20	37	15.4	173987	9	AL450337	AL450337 Human DNA
21	36.8	15.3	41101	8	SPAC144	AL132675 S.pombe c
c 22	36.6	15.2	6134	6	AX251318	AX251318 Sequence
c 23	36.6	15.2	101213	2	AC113851	AC113851 Rattus no
24	36.6	15.2	146680	2	AC095815	AC095815 Rattus no
25	36.6	15.2	154960	2	AC011948	AC011948 Homo sapi
c 26	36.2	15.1	153735	9	AC004029	AC004029 Human BAC
c 27	36	15.0	13416	1	AE007832	AE007832 Clostridi
28	36	15.0	144864	2	AC096890	AC096890 Rattus no
c 29	36	15.0	169968	2	AC099745	AC099745 Bos taura
c 30	36	15.0	173786	2	AC107582	AC107582 Rattus no
c 31	36	15.0	191720	2	AC109982	AC109982 Rattus no
c 32	36	15.0	225407	2	AC124428	AC124428 Mus muscu
c 33	35.8	14.9	62418	2	AC104025	AC104025 Homo sapi
34	35.8	14.9	174186	2	AC025252	AC025252 Homo sapi
35	35.8	14.9	174293	2	AC068787	AC068787 Homo sapi
36	35.8	14.9	196698	9	CN505TEI	AL359220 Human chr
37	35.8	14.9	234817	10	AL663048	AL663048 Mouse DNA
38	35.8	14.9	347726	2	AC098521	AC098521 Rattus no
39	35.6	14.8	148756	2	AC119598	AC119598 Rattus no
c 40	35.6	14.8	170659	2	AC112694	AC112694 Homo sapi
41	35.6	14.8	196410	9	AC027544	AC027544 Homo sapi
c 42	35.6	14.8	198565	2	AC016750	AC016750 Homo sapi
43	35.6	14.8	255839	2	AC105842	AC105842 Rattus no
c 44	35.4	14.8	8623	3	AY082376	AY082376 DictyOste
c 45	35.4	14.8	28759	9	AC062027	AC062027 Homo sapi

ALIGNMENTS

RESULT 1
SCMRPTAA
LOCUS SCMRPTAA 121 bp DNA linear INV 26-APR-1993
DEFINITION S.mansoni tandem repeat units.
ACCESSION M61098
VERSION M61098.1 GI:161064
KEYWORDS .
SOURCE S.mansoni (strain Egyptian) DNA, clone pSm1-7.
ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE 1 (bases 1 to 121)
AUTHORS Hamburger,J., Turetski,T., Kapeller,I. and Deresiewicz,R.
TITLE Highly repeated short DNA sequences in the genome of Schistosoma mansoni recognized by a species-specific probe

```
  JOURNAL   Mol. Biochem. Parasitol. 44 (1), 73-80 (1991)
  MEDLINE   91187067
   PUBMED   2011155
FEATURES             Location/Qualifiers
     source          1. .121
                     /organism="Schistosoma mansoni"
                     /db_xref="taxon:6183"
BASE COUNT       36 a      20 c      26 g      39 t
ORIGIN

  Query Match          18.4%;  Score 44.2;  DB 3;  Length 121;
  Best Local Similarity   93.9%;  Pred. No. 0.14;
  Matches   46;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      81 CTCCGAAACCACTGGACGGATTTTATGATGTTTGTTTAGAGGCTTTG 129
           ||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db      41 CTCCGAAAACCACTGGACGGATTTTATGATGTTTGTTTTAGATTATTTG 89


RESULT 2
AC087493
LOCUS       AC087493               69028 bp    DNA     linear   PRI 20-MAR-2002
DEFINITION  Homo sapiens chromosome 17, clone RP11-253N19, complete sequence.
ACCESSION   AC087493
VERSION     AC087493.4  GI:19551146
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 69028)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-253N19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 69028)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Riebackt,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 69028)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Riebackt,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 20, 2002 this sequence version replaced gi:14192963.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: 11973
                Center clone name: 253_N_19
            --------------

            SP6 end overlaps AC025211 [WICGR project L78401] by 111469 bp. We
            are only submitting the final 69.0 kilobases (55.2 kb unique).
FEATURES             Location/Qualifiers
     source          1. .69028
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-253N19"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   1. .46
                     /rpt_family="polypyrimidine"
     repeat_region   complement(604. .910)
                     /rpt_family="AluSg"
     repeat_region   1210. .1488
                     /rpt_family="L1MC5"
     repeat_region   1493. .1785
                     /rpt_family="L1M4"
     repeat_region   complement(1786. .2113)
                     /rpt_family="MER1B"
     unsure          complement(1999. .2004)
                     /note="<30 qual SNSL region"
     repeat_region   2114. .2311
                     /rpt_family="L1M4"
     repeat_region   2312. .2621
                     /rpt_family="AluSx"
     repeat_region   2622. .3083
                     /rpt_family="L1M4"
     repeat_region   complement(3212. .3516)
                     /rpt_family="AluJo"
     repeat_region   complement(5102. .5229)
                     /rpt_family="PLAM_A"
     repeat_region   6076. .6244
                     /rpt_family="L1ME"
     repeat_region   complement(6280. .6332)
                     /rpt_family="MIR"
     repeat_region   6543. .6590
                     /rpt_family="AT_rich"
     repeat_region   complement(6743. .6870)
                     /rpt_family="MER8"
     repeat_region   complement(6914. .7212)
                     /rpt_family="AluY"
     repeat_region   7356. .7720
                     /rpt_family="THE1B"
     repeat_region   complement(7769. .8057)
                     /rpt_family="AluSx"
     repeat_region   8115. .8321
```

```
                     /rpt_family="L1M4"
     repeat_region   complement(8353. .8645)
                     /rpt_family="AluY"
     repeat_region   8690. .8924
                     /rpt_family="L1M4"
     repeat_region   9219. .9782
                     /rpt_family="MER68"
     repeat_region   9920. .10074
                     /rpt_family="HAL1b"
     repeat_region   10819. .10945
                     /rpt_family="MER94"
     repeat_region   11828. .12729
                     /rpt_family="L1MC/D"
     repeat_region   13015. .13294
                     /rpt_family="AluSg"
     repeat_region   13303. .13561
                     /rpt_family="HAL1b"
     repeat_region   13614. .13635
                     /rpt_family="AT_rich"
     repeat_region   13861. .13965
                     /rpt_family="L1M4"
     repeat_region   13986. .14416
                     /rpt_family="L1M4"
     repeat_region   complement(14420. .14685)
                     /rpt_family="MLT1E"
     repeat_region   complement(14798. .15074)
                     /rpt_family="L1M4"
     repeat_region   complement(15298. .15570)
                     /rpt_family="AluSx"
     repeat_region   complement(15698. .16199)
                     /rpt_family="L1PA10"
     repeat_region   16203. .16989
                     /rpt_family="L1PA10"
     repeat_region   17059. .17707
                     /rpt_family="L2"
     repeat_region   17735. .18157
                     /rpt_family="L2"
     repeat_region   complement(19128. .19641)
                     /rpt_family="HAL1"
     repeat_region   complement(19779. .20031)
                     /rpt_family="L1ME2"
     repeat_region   20036. .20064
                     /rpt_family="L2"
     repeat_region   20065. .20363
                     /rpt_family="AluSx"
     repeat_region   20364. .21000
                     /rpt_family="L2"
     repeat_region   complement(21001. .21299)
                     /rpt_family="AluY"
     repeat_region   21300. .21703
                     /rpt_family="L2"
     repeat_region   21726. .21757
                     /rpt_family="AT_rich"
     repeat_region   complement(21940. .22176)
                     /rpt_family="MIR"
     repeat_region   complement(22213. .22443)
                     /rpt_family="MIR"
     repeat_region   complement(22559. .22615)
                     /rpt_family="L2"
     repeat_region   23270. .23317
                     /rpt_family="MIR"
     repeat_region   complement(23894. .24372)
                     /rpt_family="MLT1D"
     repeat_region   24422. .24803
                     /rpt_family="MER21B"
     repeat_region   24809. .25060
                     /rpt_family="MER21B"
     repeat_region   25134. .25331
                     /rpt_family="MLT1H"
     repeat_region   26573. .26886
                     /rpt_family="L1MEc"
     repeat_region   26887. .27206
                     /rpt_family="AluSx"
     repeat_region   27207. .27771
                     /rpt_family="L1MEc"

  Query Match             17.0%;  Score 40.8;  DB 9;  Length 69028;
  Best Local Similarity   49.5%;  Pred. No. 0.8;
  Matches  105;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy       22 TCTATGAAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATAC 81
            | ||| | | || |  ||    || |||   | || |  | | || |  |  |||  || 
Db    11615 TATATTACACTCTTAATAAGATGATTTAATATTGTAGAAATGTTAATCATTTCAAATTAA 11674

Qy       82 TCCGAAACCACTGGACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTA 141
            ||   |  || |  |    |  | |  ||  |||  ||  |     |  ||  ||    |
Db    11675 TCTATATTCAGTTCAATCCTAATAAAAATTATTGAATTTAATAAACTTATGCACTTAAAA 11734

Qy      142 AAAATACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAAC 201
               ||||| || |  ||| |  ||| |    |    |   | |||||| | ||  |||| 
Db    11735 TGTATACTGCAGAATAAAGGTCTATGTATAGGTTAATCAACCTTAATAAAGAAGTAGAAA 11794

Qy      202 TAATAAACGCTCTCGCACCCGCAATTATATTT 233
              ||||||   | | |   | || | |||| |
Db    11795 AGATAAACTAACCCTCTAACACACTAATATAT 11826


RESULT 3
AC025211
LOCUS       AC025211               150794 bp    DNA     linear   PRI 08-SEP-2001
DEFINITION  Homo sapiens chromosome , clone RP11-220C2, complete sequence.
ACCESSION   AC025211
VERSION     AC025211.8  GI:15528921
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 150794)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome, clone RP11-220C2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 150794)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 150794)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
```

```
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meidrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,K., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,R.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 8, 2001 this sequence version replaced gi:18431159.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------------ Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIGR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
               Center project name: L7B40
               Center clone name: 220_C_2
            --------------

            SP6 end overlaps AC002119 [WICGR L14l] by 43891 bp. We will submit
            150793 bp of the overlap.
FEATURES             Location/Qualifiers
     source          1..150794
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-220C2"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(1..23)
                     /rpt_family="L1ME4A"
     unsure          6..23
                     /note="<30 qual SNGL region"
     repeat_region   24..328
                     /rpt_family="AluSc"
     repeat_region   complement(329..425)
                     /rpt_family="L1ME4A"
     repeat_region   complement(599..828)
                     /rpt_family="AluJb"
     repeat_region   complement(987..1086)
                     /rpt_family="L1ME4A"
     repeat_region   complement(1087..1381)
                     /rpt_family="AluJb"
     repeat_region   complement(1382..1430)
                     /rpt_family="L1ME4A"
     repeat_region   complement(1532..1780)
                     /rpt_family="L1ME4A"
     repeat_region   complement(1816..2132)
                     /rpt_family="AluSx"
     repeat_region   complement(3188..3312)
                     /rpt_family="L2"
     repeat_region   complement(3624..3919)
                     /rpt_family="AluSx"
     repeat_region   complement(4349..4640)
                     /rpt_family="AluSx"
     repeat_region   complement(4861..5224)
                     /rpt_family="L2"
     repeat_region   5818..5936
                     /rpt_family="L2"
     repeat_region   complement(5992..6226)
                     /rpt_family="MIR"
     repeat_region   7173..7280
                     /rpt_family="MER94"
     repeat_region   7486..7679
                     /rpt_family="L2"
     repeat_region   complement(8988..9267)
                     /rpt_family="AluY"
     repeat_region   10907..11697
                     /rpt_family="L2"
     repeat_region   complement(11992..12132)
                     /rpt_family="HAL1"
     repeat_region   complement(12133..12440)
                     /rpt_family="AluJo"
     repeat_region   complement(12441..12462)
                     /rpt_family="HAL1"
     repeat_region   complement(12463..12767)
                     /rpt_family="AluY"
     repeat_region   complement(12768..12898)
                     /rpt_family="HAL1"
     repeat_region   13422..13629
                     /rpt_family="MIR"
     repeat_region   15268..15291
                     /rpt_family="(TGAA)n"
     repeat_region   complement(15292..15580)
                     /rpt_family="AluJo"
     repeat_region   15584..15883
                     /rpt_family="AluSx"
     unsure          complement(15651..15659)
                     /note="<30 qual SNGL region"
     unsure          complement(15684..15701)
                     /note="<30 qual SNGL region"
     unsure          complement(15859..15865)
                     /note="<30 qual SNGL region"
     repeat_region   complement(15893..16047)
                     /rpt_family="AluJo"
     repeat_region   complement(16350..16654)
                     /rpt_family="AluY"
     repeat_region   17539..17654
                     /rpt_family="L2"
     repeat_region   complement(17671..17984)
                     /rpt_family="AluJb"
     repeat_region   complement(18115..18681)
                     /rpt_family="L1ME1"
     repeat_region   complement(18783..19023)
                     /rpt_family="L1ME1"
     repeat_region   complement(19024..19321)
                     /rpt_family="AluSg"
     repeat_region   complement(19322..20092)
                     /rpt_family="L1ME1"
     repeat_region   20098..20194
                     /rpt_family="CT-rich"
     repeat_region   complement(20195..20356)
                     /rpt_family="AluJb"
     repeat_region   20357..20395
                     /rpt_family="(TG)n"
     repeat_region   complement(20396..20690)
                     /rpt_family="AluSx"
     repeat_region   complement(20691..20830)
                     /rpt_family="AluJb"
     repeat_region   complement(20831..22342)
                     /rpt_family="L1ME1"
     unsure          complement(21587..21602)
                     /note="<30 qual SNGL region"
     unsure          complement(21644..21658)
                     /note="<30 qual SNGL region"
     repeat_region   complement(22343..22535)
                     /rpt_family="AluJb"
     repeat_region   complement(22629..22718)
                     /rpt_family="AluJ/FLAM"
     repeat_region   complement(23036..23116)
                     /rpt_family="L1MB2"
     repeat_region   complement(23117..23399)
                     /rpt_family="AluJo"
     unsure          23294..23299
                     /note="<30 qual SNGL region"
     repeat_region   complement(23400..23422)
```

```
                    /rpt_family="L1MB2"
     repeat_region  complement(23472. .23561)
                    /rpt_family="MIR"
     repeat_region  24063. .24374
                    /rpt_family="AluSx"
     repeat_region  complement(24989. .25192)
                    /rpt_family="L3"

  Query Match            17.0%;  Score 40.8;  DB 9;  Length 150794;
  Best Local Similarity   49.5%;  Pred. No. 0.76;
  Matches  105;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy      22 TCTATGAAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATAC 81
             | ||| | | || |  ||    || |||   | || |  | | || |  |  |||  ||
Db  148581 TATATTACACTCTTAATAAGATGATTTAATATTGTAGAAATGTTAATCATTTCAAATTAA 148640

Qy      82 TCCGAAACCACTGGACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTA 141
             ||  |  || |  |     |  | |  ||  |||  ||  |      |  ||  ||    |
Db  148641 TCTATATTCAGTTCAATCCTAATAAAAATTATTGAATTTAATAAACTTATGCACTTAAAA 148700

Qy     142 AAAATACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAAC 201
               ||||| || |  ||| |  ||| |      |     |  |||||| | ||  ||||
Db  148701 TGTATACTGCAGAATAAAGGTCTATGTATAGGTTAATCAACCTTAATAAAGAAGTAGAAA 148760

Qy     202 TAATAAACGCTCTCGCACCCGCAATTATATTT 233
               ||||||   | | |   | || | |||| |
Db  148761 AGATAAACTAACCCTCTAACACACTAATATAT 148792


RESULT 4
AC007982/c
LOCUS       AC007982             186360 bp    DNA     linear   PRI 31-MAY-2002
DEFINITION  Homo sapiens chromosome 17, clone RP11-44H5, complete sequence.
ACCESSION   AC007982
VERSION     AC007982.2  GI:21281552
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186360)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-44H5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186360)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUL-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 186360)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 31, 2002 this sequence version replaced gi:5391470.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            ------------- Project Information
                Center project name: L1066
                Center clone name: 44_H_5
            -------------.
FEATURES             Location/Qualifiers
     source          1. .186360
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-44H5"
                     /clone_lib="RPCI-11 Human Male BAC"
     unsure          496. .501
                     /note="&lt;30 qual SNGL region."
     repeat_region   505. .648
                     /rpt_family="MER5B"
     unsure          534. .539
                     /note="&lt;30 qual SNGL region."
     unsure          554. .557
                     /note="&lt;30 qual SNGL region."
     unsure          599. .607
                     /note="&lt;30 qual SNGL region."
     repeat_region   complement(923. .1126)
                     /rpt_family="L1ME4A"
     repeat_region   complement(1153. .1446)
                     /rpt_family="L2"
     repeat_region   complement(1490. .1781)
                     /rpt_family="AluJb"
     repeat_region   2152. .2455
                     /rpt_family="AluY"
     repeat_region   complement(3160. .3275)
                     /rpt_family="MLT1L"
     repeat_region   complement(3449. .3711)
                     /rpt_family="MLT1L"
     repeat_region   3718. .3940
                     /rpt_family="L3"
     repeat_region   4189. .4680
                     /rpt_family="L1MC4a"
     repeat_region   4729. .5262
                     /rpt_family="L1MC4"
     repeat_region   5267. .5430
                     /rpt_family="L3"
     repeat_region   complement(5915. .6213)
                     /rpt_family="AluY"
     repeat_region   complement(6302. .6529)
                     /rpt_family="MIR"
     repeat_region   complement(6701. .7286)
                     /rpt_family="L2"
```

```
repeat_region   7922. .8282
                /rpt_family="L1ME3"
repeat_region   complement(9647. .10091)
                /rpt_family="L2"
repeat_region   complement(10132. .10667)
                /rpt_family="L2"
repeat_region   complement(10750. .10858)
                /rpt_family="L1ME4A"
repeat_region   complement(10954. .11091)
                /rpt_family="L1ME4A"
repeat_region   11092. .11381
                /rpt_family="AluSx"
repeat_region   complement(11382. .11440)
                /rpt_family="L1ME4A"
repeat_region   11577. .11908
                /rpt_family="MER1B"
repeat_region   complement(12134. .12469)
                /rpt_family="L2"
repeat_region   12492. .12518
                /rpt_family="(TTG)n"
repeat_region   13024. .13066
                /rpt_family="AT_rich"
repeat_region   complement(13695. .14462)
                /rpt_family="L1M2"
repeat_region   complement(15040. .15341)
                /rpt_family="L1MC4"
repeat_region   complement(15149. .15342)
                /rpt_family="L1MC4"
repeat_region   15703. .15987
                /rpt_family="AluSx"
repeat_region   complement(16516. .16816)
                /rpt_family="AluJo"
repeat_region   17718. .17894
                /rpt_family="MER3"
repeat_region   complement(18367. .18468)
                /rpt_family="L1ME4A"
repeat_region   18743. .18772
                /rpt_family="AT_rich"
repeat_region   complement(18785. .19067)
                /rpt_family="AluJo"
repeat_region   19072. .19247
                /rpt_family="(TA)n"
repeat_region   complement(19250. .19658)
                /rpt_family="L1ME4A"
repeat_region   20328. .20638
                /rpt_family="AluSx"
repeat_region   21257. .21915
                /rpt_family="L2"
repeat_region   complement(22037. .22332)
                /rpt_family="AluJb"
repeat_region   complement(22333. .22676)
                /rpt_family="MER47A"
repeat_region   22679. .22851
                /rpt_family="MLT1K"
repeat_region   22865. .22930
                /rpt_family="GA-rich"
repeat_region   23250. .23429
                /rpt_family="L2"
repeat_region   24718. .25104
                /rpt_family="L2"
repeat_region   25599. .25951
                /rpt_family="L1ME"
repeat_region   26191. .27167
                /rpt_family="L1ME3"
repeat_region   27205. .27351
                /rpt_family="MIR"
repeat_region   27579. .28009
                /rpt_family="L3"
repeat_region   complement(28009. .28328)
                /rpt_family="L2"
repeat_region   complement(28412. .28480)
                /rpt_family="L2"
repeat_region   complement(28481. .28785)
                /rpt_family="AluSc"
repeat_region   complement(28786. .29090)
                /rpt_family="L2"
repeat_region   29159. .29240
                /rpt_family="L2"
repeat_region   complement(29376. .30479)
                /rpt_family="L2"
repeat_region   30920. .31226
                /rpt_family="AluY"
repeat_region   31228. .31415
                /rpt_family="AluJb"
repeat_region   31441. .31475

  Query Match             17.0%;  Score 40.8;  DB 9;  Length 186360;
  Best Local Similarity   49.5%;  Pred. No. 0.76;
  Matches  105;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy      22 TCTATGAAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATAC 81
           | ||| | | || |  ||    || |||   | ||| | | (( | | | ||| ||
Db 153551 TATATTACACTCTTAATAAGATGATTTAATATTGTAGAAATGTTAATCATTTCAAATTAA 153492

Qy      82 TCCGAAACCACTGGACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTA 141
           ||  |  || | |    | | |  || ||| || |    |  || ||    |
Db 153491 TCTATATTCAGTTCAATCCTAATAAAAATTATTGAATTTAATAAACTTATGCACTTAAAA 153432

Qy     142 AAAATACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAAC 201
           |||||  || |  ||| |  ||| |    |    |  |||||| | || ||||
Db 153431 TGTATACTGCAGAATAAAGGTCTATGTATAGGTTAATCAACCTTAATAAAGAAGTAGAAA 153372

Qy     202 TAATAAAGCGTCTCGCACCGCAAATTATATTT 233
           |||||  |  | |  |  | ||| ||||| |
Db 153371 AGATAAACTAACCCTCTAACACACTAATATAT 153340


RESULT 5
AC008106
LOCUS       AC008106               218922 bp    DNA     linear   HTG 04-SEP-2000
DEFINITION  Homo sapiens chromosome 17 clone CTD-3008D3 map 17, WORKING DRAFT
            SEQUENCE, 34 unordered pieces.
ACCESSION   AC008106
VERSION     AC008106.3  GI:9966323
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 218922)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone CTD-3008D3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 218922)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Case,A., Carny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., DePayre,R., DeVon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychalecky,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,R., Yo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUL-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 4, 2000 this sequence version replaced gi:7329486.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
```

```
------------ Genome Center
     Center: Whitehead Institute/ MIT Center for Genome Research
     Center code: WIBR
     Web site: http://www-seq.wi.mit.edu
     Contact: sequence_submissions@genome.wi.mit.edu
------------ Project Information
     Center project name: L900
     Center clone name: 3008_D_3
------------ Summary Statistics
     Sequencing vector: M13; M77815; 100% of reads
     Chemistry: Dye-primer-amersham; 100% of reads
     Assembly program: Phrap; version 0.960731
     Consensus quality: 197177 bases at least Q40
     Consensus quality: 206822 bases at least Q30
     Consensus quality: 211643 bases at least Q20
     Insert size: 200000; agarose-fp
     Insert size: 215622; sum-of-contigs
     Quality coverage: 4.1 in Q20 bases; agarose-fp
     Quality coverage: 3.8 in Q20 bases; sum-of-contigs
------------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 34 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*       1     1090: contig of 1090 bp in length
*    1091  1190: gap of      100 bp
*    1191     2690: contig of 1500 bp in length
*    2691  2790: gap of      100 bp
*    2791     4012: contig of 1222 bp in length
*    4013  4112: gap of      100 bp
*    4113     5123: contig of 1011 bp in length
*    5124  5223: gap of      100 bp
*    5224     6312: contig of 1089 bp in length
*    6313  6412: gap of      100 bp
*    6413     8474: contig of 2062 bp in length
*    8475  8574: gap of      100 bp
*    8575    10328: contig of 1754 bp in length
*   10329 10428: gap of      100 bp
*   10429    12116: contig of 1688 bp in length
*   12117 12216: gap of      100 bp
*   12217    14183: contig of 1967 bp in length
*   14184 14283: gap of      100 bp
*   14284    17330: contig of 3047 bp in length
*   17331 17430: gap of      100 bp
*   17431    21134: contig of 3704 bp in length
*   21135 21234: gap of      100 bp
*   21235    24428: contig of 3194 bp in length
*   24429 24528: gap of      100 bp
*   24529    27705: contig of 3177 bp in length
*   27706 27805: gap of      100 bp
*   27806    30106: contig of 2301 bp in length
*   30107 30206: gap of      100 bp
*   30207    35366: contig of 5160 bp in length
*   35367 35466: gap of      100 bp
*   35467    39634: contig of 4168 bp in length
*   39635 39734: gap of      100 bp
*   39735    43767: contig of 4033 bp in length
*   43768 43867: gap of      100 bp
*   43868    50104: contig of 6237 bp in length
*   50105 50204: gap of      100 bp
*   50205    54896: contig of 4692 bp in length
*   54897 54996: gap of      100 bp
*   54997    61677: contig of 6681 bp in length
*   61678 61777: gap of      100 bp
*   61778    68042: contig of 6265 bp in length
*   68043 68142: gap of      100 bp
*   68143    76003: contig of 7861 bp in length
*   76004 76103: gap of      100 bp
*   76104    83935: contig of 7832 bp in length
*   83936 84035: gap of      100 bp
*   84036    92731: contig of 8696 bp in length
*   92732 92831: gap of      100 bp
*   92832   102205: contig of 9374 bp in length
*  102206 102305: gap of      100 bp
*  102306   111555: contig of 9250 bp in length
*  111556 111655: gap of      100 bp
*  111656   120772: contig of 9117 bp in length
*  120773 120872: gap of      100 bp
*  120873   129869: contig of 8997 bp in length
*  129870 129969: gap of      100 bp
*  129970   143160: contig of 13191 bp in length
*  143161 143260: gap of      100 bp
*  143261   154541: contig of 11281 bp in length
*  154542 154641: gap of      100 bp
*  154642   170485: contig of 15844 bp in length
*  170486 170585: gap of      100 bp
*  170586   183973: contig of 13388 bp in length
*  183974 184073: gap of      100 bp
*  184074   200761: contig of 16688 bp in length
*  200762 200861: gap of      100 bp
*  200862   218922: contig of 18061 bp in length.
FEATURES             Location/Qualifiers
     source          1. .218922
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="CTD-3008D3"
                     /clone_lib="CITD Human BAC"
     misc_feature    1. .1090
                     /note="assembly_fragment"
     misc_feature    1191. .2690
                     /note="assembly_fragment"
     misc_feature    2791. .4012
                     /note="assembly_fragment"
     misc_feature    4113. .5123
                     /note="assembly_fragment"
     misc_feature    5224. .6312
                     /note="assembly_fragment"
     misc_feature    6413. .8474
                     /note="assembly_fragment"
     misc_feature    8575. .10328
                     /note="assembly_fragment"
     misc_feature    10429. .12116
                     /note="assembly_fragment"
     misc_feature    12217. .14183
                     /note="assembly_fragment"
     misc_feature    14284. .17330
                     /note="assembly_fragment"
     misc_feature    17431. .21134
                     /note="assembly_fragment"
     misc_feature    21235. .24428
                     /note="assembly_fragment"
     misc_feature    24529. .27705
                     /note="assembly_fragment"
     misc_feature    27806. .30106
                     /note="assembly_fragment"
     misc_feature    30207. .35366
                     /note="assembly_fragment"
     misc_feature    35467. .39634
                     /note="assembly_fragment"
     misc_feature    39735. .43767
                     /note="assembly_fragment"
     misc_feature    43868. .50104
                     /note="assembly_fragment"
     misc_feature    50205. .54896
                     /note="assembly_fragment"
     misc_feature    54997. .61677
                     /note="assembly_fragment"
     misc_feature    61778. .68042
                     /note="assembly_fragment"
     misc_feature    68143. .76003
```

```
                     /note="assembly_fragment"
     misc_feature    76104..83935
                     /note="assembly_fragment"
     misc_feature    84036..92731
                     /note="assembly_fragment"
     misc_feature    92832..102205
                     /note="assembly_fragment"
     misc_feature    102306..111555
                     /note="assembly_fragment"
     misc_feature    111656..120772
                     /note="assembly_fragment"
     misc_feature    120873..129869
                     /note="assembly_fragment"
     misc_feature    129970..143160
                     /note="assembly_fragment"
     misc_feature    143261..154541
                     /note="assembly_fragment"

  Query Match             17.0%;  Score 40.8;  DB 2;  Length 218922;
  Best Local Similarity   49.5%;  Pred. No. 0.75;
  Matches  105;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy       22 TCTATGAAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATAC 81
            | ||| | | || |  ||    || |||    | || |  |  | || |  |  |||  || 
Db   194649 TATATTACACTCTTAATAAGATGATTTAATATTGTAGAAATGTTAATCATTTCAAATTAA 194708

Qy       82 TCCGAAACCACTGGACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTA 141
            ||   |  || |  |    |  | |  ||  |||  ||  |     |  ||  ||    |
Db   194709 TCTATATTCAGTTCAATCCTAATAAAAATTATTGAATTTAATAAACTTATGCACTTAAAA 194768

Qy      142 AAAATACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAAC 201
               ||||| || |  ||| |  ||| |    |    |   | |||||| | ||  |||| 
Db   194769 TGTATACTGCAGAATAAAGGTCTATGTATAGGTTAATCAACCTTAATAAAGAAGTAGAAA 194828

Qy      202 TAATAAACGCTCTCGCACCCGCAATTATATTT 233
              ||||||   | | |   | || | ||||  |
Db   194829 AGATAAACTAACCCTCTAACACACTAATATAT 194860


RESULT 6
AC095286
LOCUS       AC095286              171744 bp    DNA     linear   HTG 11-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-12C16, *** SEQUENCING IN PROGRESS
            ***, 68 unordered pieces.
ACCESSION   AC095286
VERSION     AC095286.3  GI:21723052
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 171744)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171744)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 171744)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 10, 2002 this sequence version replaced gi:17943590.
            -------------- Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
               Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
               Center project name: GDYU
               Center clone name: CH230-12C16
            -------------- Summary Statistics
               Sequencing vector: Plasmid;
               Chemistry: Dye-terminator Big Dye: 100% of reads
               Assembly program: Phrap; version 0.990329
               Consensus quality: 116672 bases at least Q40
               Consensus quality: 123027 bases at least Q30
               Consensus quality: 127822 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 68 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *          1     1396: contig of 1396 bp in length
            *       1397     1496: gap of unknown length
            *       1497     2649: contig of 1153 bp in length
            *       2650     2749: gap of unknown length
            *       2750     3906: contig of 1157 bp in length
            *       3907     4006: gap of unknown length
            *       4007     5016: contig of 1010 bp in length
            *       5017     5116: gap of unknown length
            *       5117     6328: contig of 1212 bp in length
            *       6329     6428: gap of unknown length
            *       6429     7854: contig of 1426 bp in length
            *       7855     7954: gap of unknown length
```

```
*     7955     9302: contig of 1348 bp in length
*     9303     9402: gap of unknown length
*     9403    10892: contig of 1490 bp in length
*    10893    10992: gap of unknown length
*    10993    12542: contig of 1550 bp in length
*    12543    12642: gap of unknown length
*    12643    13700: contig of 1058 bp in length
*    13701    13800: gap of unknown length
*    13801    14835: contig of 1035 bp in length
*    14836    14935: gap of unknown length
*    14936    16473: contig of 1538 bp in length
*    16474    16573: gap of unknown length
*    16574    18097: contig of 1524 bp in length
*    18098    18197: gap of unknown length
*    18198    19476: contig of 1279 bp in length
*    19477    19576: gap of unknown length
*    19577    20672: contig of 1096 bp in length
*    20673    20772: gap of unknown length
*    20773    22484: contig of 1712 bp in length
*    22485    22584: gap of unknown length
*    22585    24397: contig of 1813 bp in length
*    24398    24497: gap of unknown length
*    24498    26493: contig of 1996 bp in length
*    26494    26593: gap of unknown length
*    26594    28669: contig of 2076 bp in length
*    28670    28769: gap of unknown length
*    28770    29896: contig of 1127 bp in length
*    29897    29996: gap of unknown length
*    29997    31368: contig of 1372 bp in length
*    31369    31468: gap of unknown length
*    31469    33202: contig of 1734 bp in length
*    33203    33302: gap of unknown length
*    33303    35107: contig of 1805 bp in length
*    35108    35207: gap of unknown length
*    35208    37039: contig of 1832 bp in length
*    37040    37139: gap of unknown length
*    37140    39050: contig of 1911 bp in length
*    39051    39150: gap of unknown length
*    39151    40878: contig of 1728 bp in length
*    40879    40978: gap of unknown length
*    40979    43526: contig of 2548 bp in length
*    43527    43626: gap of unknown length
*    43627    44913: contig of 1287 bp in length
*    44914    45013: gap of unknown length
*    45014    46466: contig of 1453 bp in length
*    46467    46566: gap of unknown length
*    46567    48405: contig of 1839 bp in length
*    48406    48505: gap of unknown length
*    48506    50475: contig of 1970 bp in length
*    50476    50575: gap of unknown length
*    50576    51954: contig of 1379 bp in length
*    51955    52054: gap of unknown length
*    52055    54498: contig of 2444 bp in length
*    54499    54598: gap of unknown length
*    54599    57961: contig of 3363 bp in length
*    57962    58061: gap of unknown length
*    58062    59731: contig of 1670 bp in length
*    59732    59831: gap of unknown length
*    59832    61144: contig of 1313 bp in length
*    61145    61244: gap of unknown length
*    61245    62564: contig of 1320 bp in length
*    62565    62664: gap of unknown length
*    62665    64872: contig of 2208 bp in length
*    64873    64972: gap of unknown length
*    64973    67564: contig of 2592 bp in length
*    67565    67664: gap of unknown length
*    67665    69460: contig of 1796 bp in length
*    69461    69560: gap of unknown length
*    69561    72505: contig of 2945 bp in length
*    72506    72605: gap of unknown length
*    72606    76677: contig of 4072 bp in length
*    76678    76777: gap of unknown length
*    76778    79723: contig of 2946 bp in length
*    79724    79823: gap of unknown length
*    79824    82619: contig of 2796 bp in length
*    82620    82719: gap of unknown length
*    82720    85163: contig of 2444 bp in length
*    85164    85263: gap of unknown length
*    85264    88694: contig of 3431 bp in length
*    88695    88794: gap of unknown length
*    88795    91392: contig of 2598 bp in length
*    91393    91492: gap of unknown length
*    91493    94703: contig of 3211 bp in length
*    94704    94803: gap of unknown length
*    94804    97009: contig of 2206 bp in length
*    97010    97109: gap of unknown length
*    97110    99202: contig of 2093 bp in length
*    99203    99302: gap of unknown length
*    99303   101344: contig of 2042 bp in length
*   101345   101444: gap of unknown length
*   101445   104544: contig of 3100 bp in length
*   104545   104644: gap of unknown length
*   104645   108315: contig of 3671 bp in length
*   108316   108415: gap of unknown length
*   108416   111355: contig of 2940 bp in length

 Query Match              16.9%;  Score 40.6;  DB 2;  Length 171744;
 Best Local Similarity    62.9%;  Pred. No. 0.86;
 Matches   61;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy    141 AAAAATACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAA 200
          |||||| | ||| |||| || |||| | | || ||| | | | | | | |||||
Db   3376 AAAAAAAAAAAAAAATAAATATTTCTTTGGGGGGTCGGGGGGGGAAAGRAAAGRGAAGRA 3435

Qy    201 CTAATAAACGCCTCTCGCACCGCAATTATATTTTGTT 237
          || | | |||| |||  |||    || | || | | |
Db   3436 AAAAGACAAGCTCCCGCCCCCGNNTTGTGTTGTGGT 3472

RESULT 7
AC080073/c
LOCUS       AC080073               57075 bp    DNA     linear   PRI 09-JAN-2002
DEFINITION  Homo sapiens BAC clone RP11-321C7 from 7, complete sequence.
ACCESSION   AC080073
VERSION     AC080073.6  GI:15638817
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 57075)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
  PUBMED    9847074
REFERENCE   2  (bases 1 to 57075)
  AUTHORS   Trani,L., Koslowicz,A. and Elliott,G.
  TITLE     The sequence of Homo sapiens BAC clone RP11-321C7
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 57075)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 57075)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-SEP-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 57075)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Department of Genetics, Washington
```

```
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Sep 18, 2001 this sequence version replaced gi:15487471.
            ------------ Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            ------------ Summary Statistics
                Center project name: H_NH0321C07
            ------------.

            NOTICE: This sequence may not represent the entire insert of this
            clone. It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center. For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR: pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP4-736H5, 2000 bp overlap; the
            clone sequenced to the right is RP11-421N10, 2000 bp overlap.
            Actual start of this clone is at base position 109513 of RP4-736H5.

            Polymorphisms have been indentified between AC005482 and AC080073.
FEATURES            Location/qualifiers
     source         1..57075
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="7"
                    /map="7"
                    /clone="RP11-321C7"
                    /clone_lib="RPCI-11"
     repeat_region  2..68
                    /rpt_family="MIR"
     repeat_region  69..431
                    /rpt_family="MaLR"
     repeat_region  500..774
                    /rpt_family="L2"
     repeat_region  1485..1589
                    /rpt_family="L2"
     repeat_region  2070..2443
                    /rpt_family="MaLR"
     repeat_region  3326..3621
                    /rpt_family="Alu"
     repeat_region  3623..3742
                    /rpt_family="MER1_type"
     repeat_region  3842..3956
                    /rpt_family="MIR"
     repeat_region  4038..4338
                    /rpt_family="Alu"
     repeat_region  4342..4499
                    /rpt_family="Alu"
     repeat_region  4498..4535
                    /rpt_family="(TAA)n"
     repeat_region  4534..4608
                    /rpt_family="ERV1"
     repeat_region  4744..4780
                    /rpt_family="(TTTTTG)n"
     repeat_region  5396..5707
                    /rpt_family="Alu"
     repeat_region  5798..6190
                    /rpt_family="ERV1"
     repeat_region  6611..6786
                    /rpt_family="Alu"
     repeat_region  6798..6847
                    /rpt_family="Alu"
     repeat_region  7496..7594
                    /rpt_family="ERV1"
     repeat_region  7591..7966
                    /rpt_family="ERV1"
     repeat_region  8041..8340
                    /rpt_family="Alu"
     repeat_region  8342..8738
                    /rpt_family="ERVL"
     repeat_region  8697..8720
                    /rpt_family="AT_rich"
     repeat_region  8799..9087
                    /rpt_family="Alu"
     repeat_region  8949..8969
                    /rpt_family="AT_rich"
     repeat_region  9154..9227
                    /rpt_family="L2"
     repeat_region  9228..9848
                    /rpt_family="ERV1"
     repeat_region  9849..9917
                    /rpt_family="L2"
     repeat_region  10095..10828
                    /rpt_family="ERV1"
     repeat_region  10955..11054
                    /rpt_family="L2"
     repeat_region  11055..11355
                    /rpt_family="Alu"
     repeat_region  11356..11490
                    /rpt_family="L2"
     repeat_region  11832..12581
                    /rpt_family="ERV1"
     repeat_region  13253..13552
                    /rpt_family="Alu"
     repeat_region  13370..13397
                    /rpt_family="AT_rich"
     repeat_region  13611..14247
                    /rpt_family="ERV1"
     repeat_region  14248..14810
                    /rpt_family="ERV1"
     repeat_region  14887..15499
                    /rpt_family="L1"
     repeat_region  15504..16135
                    /rpt_family="ERV1"
     repeat_region  16136..16218
                    /rpt_family="ERV1"
     repeat_region  16682..16837
                    /rpt_family="ERV1"
     repeat_region  16851..16876
                    /rpt_family="AT_rich"
     repeat_region  16857..17162
                    /rpt_family="Alu"
     repeat_region  17178..17665
                    /rpt_family="L1"
     repeat_region  17666..17919
                    /rpt_family="ERV1"
```

```
     repeat_region   17920. .18235
                     /rpt_family="Alu"
     repeat_region   18236. .18347
                     /rpt_family="ERV1"
     repeat_region   18503. .18614
                     /rpt_family="ERV1"
     repeat_region   18770. .19042
                     /rpt_family="L1"
     repeat_region   19060. .19364
                     /rpt_family="L1"
     repeat_region   19376. .19671
                     /rpt_family="Alu"
     repeat_region   19495. .19517
                     /rpt_family="AT_rich"
     repeat_region   19675. .19842
                     /rpt_family="L1"
     repeat_region   19689. .19766
                     /rpt_family="(TA)n"
     repeat_region   19767. .19820

  Query Match             16.6%;  Score 39.8;  DB 9;  Length 57075;
  Best Local Similarity   60.7%;  Pred. No. 1.5;
  Matches   65;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy    103 TTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAACAAAATGA 162
            || |||| ||||||||    |   ||||  |     |  | |||||  |   | ||| ||| |
Db  35441 TTGATGAAGTTTGTTTTCTTGTTTTTGTTTTTTTTCAAAAAAAAAAAAAAAAAAACTCC 35382

Qy    163 TTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAACTAATAAAC 209
                | |||| || |||  |   |   || |  |||  |   ||| || ||||
Db  35381 AGCCTGGCGACAGAGTGAGACTCCATCTCAAAAACAAACAAACAAAC 35335


RESULT 8
DJ526N18/c
LOCUS       DJ526N18             180104 bp    DNA     linear   PRI 09-FEB-1999
DEFINITION  Homo sapiens BAC526N18 neurexin III gene, partial cds.
ACCESSION   AF123462
VERSION     AF123462.1  GI:4240565
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180104)
  AUTHORS   Multimegabase Sequencing Group.
  TITLE     Sequence of neurexin III gene in human chromosome 14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180104)
  AUTHORS   Multimegabase Sequencing Group.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-1999) Department of Molecular Biotechnology, Box
            357730 University of Washington, Seattle, Washington 98195, USA
COMMENT     Sequencing methodology:  high redundancy shotgun in plasmids.
            Interspersed Repeats were identified with RepeatMasker (available
            from http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES             Location/Qualifiers
     source          1. .180104
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14q24.3-31"
                     /clone="BAC526N18"
                     /clone_lib="P. deJong RPCI11 library"
     CDS             join(<546. .624,56473. .57398)
                     /note="Intron-exon boundaries defined in relation to mRNA
                     KIAA0743, found in GenBank Accession Number AB018286."
                     /codon_start=1
                     /product="neurexin III"
                     /protein_id="AAD13621.1"
                     /db_xref="GI:4240566"
                     /translation="PTSDDLVSSAECSSDDEDFVECEPSTGGELVIPLLVEDPLATPP
                     IATRAPSITLPPTFRPLLTIIETTKDSLSMTSEAGLPCLSDQGSDGCDDDGLVISGYG
                     SGETFDSNLPPTDDEDFYTTFSLVTDKSLSTSIFEGGYKAHAPKWESKDFRPNKVSET
                     SRTTTTSLSPELIRFTASSSSGMVPKLPAGKMNNRDLKPQPDIVLLPLPTAYELDSTK
                     LKSPLITSPMFRNVPTANPTEPGIRRVPGASEVIRESSSTTGMVVGIVAAAALCILIL
                     LYAMYKYRNRDEGSYQVDETRNYISNSAQSNGTLMKEKQQSSKSGHKKQKNKDREYYV
                     "
     repeat_region   679. .790
                     /rpt_family="(TA)n"
     repeat_region   complement(2076. .2106)
                     /rpt_family="LINE2"
     repeat_region   2373. .2397
                     /rpt_family="(TA)n"
     repeat_region   complement(2410. .2671)
                     /rpt_family="AluJo"
     unsure          2470. .2480
                     /note="low quality data"
     repeat_region   3459. .3690
                     /rpt_family="AluSg1"
     repeat_region   complement(3691. .3729)
                     /rpt_family="(CATA)n"
     repeat_region   4215. .4242
                     /rpt_family="AT_rich"
     repeat_region   complement(4670. .4714)
                     /rpt_family="AT_rich"
     repeat_region   complement(4727. .4872)
                     /rpt_family="AluJb"
     repeat_region   complement(7014. .7315)
                     /rpt_family="AluY"
     repeat_region   7545. .7578
                     /rpt_family="AT_rich"
     repeat_region   8414. .8504
                     /rpt_family="(TA)n"
     repeat_region   complement(8735. .8806)
                     /rpt_family="(TA)n"
     repeat_region   10977. .11055
                     /rpt_family="(GAAA)n"
     repeat_region   11084. .11480
                     /rpt_family="MER63B"
     repeat_region   complement(11482. .11655)
                     /rpt_family="MIR"
     repeat_region   complement(12496. .12786)
                     /rpt_family="AluSx"
     repeat_region   complement(12936. .13146)
                     /rpt_family="MIR"
     repeat_region   complement(13488. .13786)
                     /rpt_family="AluY"
     repeat_region   14552. .14759
                     /rpt_family="MER58A"
     repeat_region   complement(14917. .15216)
                     /rpt_family="AluSg"
     repeat_region   complement(15955. .16187)
                     /rpt_family="L1PA14"
     repeat_region   complement(17103. .17137)
                     /rpt_family="POLY_A"
     repeat_region   complement(17213. .17300)
                     /rpt_family="MIR"
     repeat_region   complement(17325. .17360)
                     /rpt_family="AT_rich"
     repeat_region   18242. .18281
                     /rpt_family="(CAAA)n"
     repeat_region   18856. .19028
                     /rpt_family="AluJb"
     repeat_region   complement(20629. .20660)
                     /rpt_family="POLY_A"
     repeat_region   complement(21435. .21726)
                     /rpt_family="AluSx"
     repeat_region   complement(21751. .22193)
                     /rpt_family="L1MC4"
     repeat_region   complement(22480. .22508)
                     /rpt_family="MIR"
     repeat_region   complement(22495. .22632)
                     /rpt_family="LINE2"
     repeat_region   complement(22904. .23125)
                     /rpt_family="MSTA"
```

```
repeat_region   complement(23126. .23429)
                /rpt_family="AluSq"
repeat_region   complement(23430. .23595)
                /rpt_family="MSVA"
repeat_region   23772. .24072
                /rpt_family="AluSx"
repeat_region   27422. .27598
                /rpt_family="MER2"
repeat_region   27906. .28113
                /rpt_family="MER20"
unsure          28100. .28200
                /note="low quality data"
repeat_region   complement(28333. .28353)
                /rpt_family="AT_rich"
repeat_region   complement(30347. .30515)
                /rpt_family="L1MA6"
repeat_region   30499. .30755
                /rpt_family="L1MA8"
repeat_region   complement(34668. .34742)
                /rpt_family="MER57_internal"
repeat_region   complement(34901. .35194)
                /rpt_family="L1"
repeat_region   complement(35391. .35448)
                /rpt_family="(CATA)n"
repeat_region   complement(37534. .37819)
                /rpt_family="L1M2"
repeat_region   complement(38029. .38595)
                /rpt_family="L1MA3"
repeat_region   complement(38601. .39187)
                /rpt_family="TIGGER1"
repeat_region   complement(39198. .39562)
                /rpt_family="L1PB2"
repeat_region   complement(39565. .40986)
                /rpt_family="TIGGER1"
repeat_region   40994. .41121
                /rpt_family="(GA)n"
repeat_region   41124. .41217
                /rpt_family="(GA)n"
repeat_region   41222. .41361
                /rpt_family="(GA)n"
unsure          41280. .41450
                /note="low quality data"
repeat_region   complement(41379. .41710)
                /rpt_family="TIGGER1"
repeat_region   complement(41717. .41738)
                /rpt_family="AT_rich"
repeat_region   complement(43779. .43994)
                /rpt_family="MER30"
repeat_region   44206. .44701
                /rpt_family="L1ME2"
repeat_region   44702. .44948
                /rpt_family="L1"
repeat_region   44801. .45256
                /rpt_family="L1MA4"
repeat_region   complement(46920. .46948)
                /rpt_family="AT_rich"
unsure          47640. .47740
                /note="low quality data"
repeat_region   47669. .47714
                /rpt_family="(CA)n"
repeat_region   48290. .48325
                /rpt_family="(TA)n"
repeat_region   complement(49741. .49871)
                /rpt_family="(TAAAA)n"
repeat_region   complement(50447. .50501)
                /rpt_family="(GA)n"
repeat_region   50983. .51063
                /rpt_family="AT_rich"
repeat_region   complement(52039. .52067)
                /rpt_family="POLY_A"
repeat_region   complement(52656. .52955)
                /rpt_family="AluY"
repeat_region   complement(53075. .53096)
                /rpt_family="AT_rich"
repeat_region   57850. .57933
                /rpt_family="POLY_A"
repeat_region   58688. .58779
                /rpt_family="(TAAA)n"
repeat_region   complement(58742. .58783)
                /rpt_family="(TA)n"
repeat_region   complement(60184. .60483)
                /rpt_family="AluY"
repeat_region   complement(60629. .60652)
                /rpt_family="AT_rich"
repeat_region   60904. .60924
                /rpt_family="AT_rich"
repeat_region   complement(61082. .61360)
                /rpt_family="AluSx"
repeat_region   complement(62520. .62627)
```

```
 Query Match             16.3%;  Score 39.2;  DB 9;  Length 180104;
 Best Local Similarity   60.2%;  Pred. No. 2;
 Matches   65;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy      98 GGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAACAA 157
           || || || |  | ||| ||||| ||   ||||| ||||   |  |||  || | | ||
Db 177891 GGCTTCTTGTTTTTTTTCTTTTACAGAAATTGGTAACCTTGTTCTAAAATTTATACAGAA 177832

Qy     158 AATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAACTAAT 205
           |  |  | ||| || || |   || | ||| |||||| |||
Db 177831 AGGCAGGAAATCCGAAAGAGTAGAAATTATTTAAAAAAGAACAAAT 177784


RESULT 9
AC109963/c
LOCUS       AC109963              166434 bp    DNA     linear   HTG 13-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-139E1, *** SEQUENCING IN PROGRESS
            ***, 82 unordered pieces.
ACCESSION   AC109963
VERSION     AC109963.3  GI:21738204
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 166434)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowles,S., Brieva,M., Brown,S., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Duqah-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Holmes,S.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,C., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,N., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
```

```
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan, I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,R.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 166434)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 166434)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 12, 2002 this sequence version replaced gi:18860231.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GQLI
                Center clone name: CH230-139E1
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 66236 bases at least Q40
                Consensus quality: 73589 bases at least Q30
                Consensus quality: 80243 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 82 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1080: contig of 1080 bp in length
            *    1081     1190: gap of unknown length
            *    1181     2319: contig of 1139 bp in length
            *    2320     2419: gap of unknown length
            *    2420     3571: contig of 1152 bp in length
            *    3572     3671: gap of unknown length
            *    3672     4958: contig of 1287 bp in length
            *    4959     5058: gap of unknown length
            *    5059     6429: contig of 1371 bp in length
            *    6430     6529: gap of unknown length
            *    6530     7724: contig of 1195 bp in length
            *    7725     7824: gap of unknown length
            *    7825     8884: contig of 1060 bp in length
            *    8885     8984: gap of unknown length
            *    8985    10255: contig of 1271 bp in length
            *   10256    10355: gap of unknown length
            *   10356    11657: contig of 1302 bp in length
            *   11658    11757: gap of unknown length
            *   11758    12879: contig of 1122 bp in length
            *   12880    12979: gap of unknown length
            *   12980    14104: contig of 1125 bp in length
            *   14105    14204: gap of unknown length
            *   14205    15282: contig of 1078 bp in length
            *   15283    15382: gap of unknown length
            *   15383    16598: contig of 1216 bp in length
            *   16599    16698: gap of unknown length
            *   16699    18184: contig of 1486 bp in length
            *   18185    18284: gap of unknown length
            *   18285    19571: contig of 1287 bp in length
            *   19572    19671: gap of unknown length
            *   19672    20903: contig of 1232 bp in length
            *   20904    21003: gap of unknown length
            *   21004    22092: contig of 1089 bp in length
            *   22093    22192: gap of unknown length
            *   22193    23299: contig of 1107 bp in length
            *   23300    23399: gap of unknown length
            *   23400    24504: contig of 1105 bp in length
            *   24505    24604: gap of unknown length
            *   24605    25806: contig of 1202 bp in length
            *   25807    25906: gap of unknown length
            *   25907    26993: contig of 1087 bp in length
            *   26994    27093: gap of unknown length
            *   27094    28509: contig of 1416 bp in length
            *   28510    28609: gap of unknown length
            *   28610    29874: contig of 1265 bp in length
            *   29875    29974: gap of unknown length
            *   29975    31229: contig of 1255 bp in length
            *   31230    31329: gap of unknown length
            *   31330    32613: contig of 1284 bp in length
            *   32614    32713: gap of unknown length
            *   32714    33896: contig of 1183 bp in length
            *   33897    33996: gap of unknown length
            *   33997    35668: contig of 1672 bp in length
            *   35669    35768: gap of unknown length
            *   35769    36866: contig of 1098 bp in length
            *   36867    36966: gap of unknown length
            *   36967    38714: contig of 1748 bp in length
            *   38715    38814: gap of unknown length
            *   38815    40155: contig of 1341 bp in length
            *   40156    40255: gap of unknown length
            *   40256    41348: contig of 1093 bp in length
            *   41349    41448: gap of unknown length
            *   41449    43123: contig of 1675 bp in length
            *   43124    43223: gap of unknown length
            *   43224    44800: contig of 1577 bp in length
            *   44801    44900: gap of unknown length
            *   44901    46797: contig of 1897 bp in length
            *   46798    46897: gap of unknown length
            *   46898    48428: contig of 1531 bp in length
            *   48429    48528: gap of unknown length
            *   48529    49909: contig of 1381 bp in length
            *   49910    50009: gap of unknown length
            *   50010    51370: contig of 1361 bp in length
            *   51371    51470: gap of unknown length
            *   51471    52938: contig of 1468 bp in length
            *   52939    53038: gap of unknown length
            *   53039    54229: contig of 1191 bp in length
            *   54230    54329: gap of unknown length
            *   54330    56173: contig of 1844 bp in length
            *   56174    56273: gap of unknown length
            *   56274    57450: contig of 1177 bp in length
            *   57451    57550: gap of unknown length
            *   57551    58689: contig of 1139 bp in length
            *   58690    58789: gap of unknown length
            *   58790    60423: contig of 1634 bp in length
            *   60424    60523: gap of unknown length
            *   60524    61794: contig of 1271 bp in length
            *   61795    61894: gap of unknown length
            *   61895    63284: contig of 1390 bp in length
            *   63285    63384: gap of unknown length
            *   63385    64680: contig of 1296 bp in length
            *   64681    64780: gap of unknown length
            *   64781    66023: contig of 1243 bp in length
            *   66024    66123: gap of unknown length
```

```
                *    66124    67877: contig of 1754 bp in length
                *    67878    67977: gap of unknown length
                *    67978    70617: contig of 2640 bp in length
                *    70618    70717: gap of unknown length
                *    70718    72859: contig of 2142 bp in length
                *    72860    72959: gap of unknown length
                *    72960    74745: contig of 1786 bp in length
                *    74746    74845: gap of unknown length
                *    74846    77563: contig of 2718 bp in length
                *    77564    77663: gap of unknown length
                *    77664    79349: contig of 1686 bp in length
                *    79350    79449: gap of unknown length
                *    79450    81055: contig of 1606 bp in length

  Query Match          16.0%;  Score 38.4;  DB 2;  Length 166434;
  Best Local Similarity   56.2%;  Pred. No. 3.4;
  Matches   72;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy      94 GGACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAA 153
           ||  ||  ||||  |  | ||| ||||   || |    ||  |  ||||| |  | ||
Db  11077 GGGGGGGGTTTTTTTTTTTTTTTTTTCCCTCCCACCCCCCAAAAAAAAAAAAAA 11018

Qy     154 ACAAAATGATTATTTGCGAGAGCGTGGCGTTAATATAAAACAAGAACTAATAAACGCTC 213
           | || |   || |||     | | | |||  |  || | ||||  ||  ||  || |||| | |
Db  11017 AAAATAATTTTTTTTTTGGGGGGGGGGGAAAAGAAAAAAAAAAAAAAAAAAACCCCC 10958

Qy     214 TCGCACCC 221
           | ||| ||||
Db  10957 CCCCCCCC 10950


RESULT 10
AC095187/c
LOCUS       AC095187               192153 bp    DNA     linear   HTG 11-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-9C3, *** SEQUENCING IN PROGRESS ***,
            79 unordered pieces.
ACCESSION   AC095187
VERSION     AC095187.3  GI:21722694
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 192153)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,N.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,L., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.P., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.I., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denu,A.L., Ding,T., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,T., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Perers,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,R., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villaion,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 192153)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 192153)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 10, 2002 this sequence version replaced gi:17942286.
            -------------- Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
               Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
               Center project name: GDBX
               Center clone name: CH230-9C3
            -------------- Summary Statistics
               Sequencing vector: Plasmid;
               Chemistry: dye-terminator Big Dye; 100% of reads
               Assembly program: Phrap; version 0.990329
               Consensus quality: 119125 bases at least Q40
               Consensus quality: 125091 bases at least Q30
               Consensus quality: 130301 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 79 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1243: contig of 1243 bp in length
            *    1244     1343: gap of unknown length
            *    1344     2721: contig of 1378 bp in length
            *    2722     2821: gap of unknown length
            *    2822     3941: contig of 1120 bp in length
            *    3942     4041: gap of unknown length
            *    4042     5289: contig of 1248 bp in length
            *    5290     5389: gap of unknown length
            *    5390     6956: contig of 1567 bp in length
            *    6957     7056: gap of unknown length
            *    7057     8133: contig of 1077 bp in length
            *    8134     8233: gap of unknown length
            *    8234     9265: contig of 1032 bp in length
            *    9266     9365: gap of unknown length
            *    9366    10944: contig of 1579 bp in length
            *   10945    11044: gap of unknown length
            *   11045    12750: contig of 1706 bp in length
            *   12751    12850: gap of unknown length
            *   12851    13958: contig of 1108 bp in length
            *   13959    14058: gap of unknown length
```

```
*    14059    15275: contig of 1217 bp in length
*    15276    15375: gap of unknown length
*    15376    16793: contig of 1418 bp in length
*    16794    16893: gap of unknown length
*    16894    18019: contig of 1126 bp in length
*    18020    18119: gap of unknown length
*    18120    19768: contig of 1649 bp in length
*    19769    19868: gap of unknown length
*    19869    21147: contig of 1279 bp in length
*    21148    21247: gap of unknown length
*    21248    23000: contig of 1753 bp in length
*    23001    23100: gap of unknown length
*    23101    25073: contig of 1973 bp in length
*    25074    25173: gap of unknown length
*    25174    27094: contig of 1921 bp in length
*    27095    27194: gap of unknown length
*    27195    28757: contig of 1563 bp in length
*    28758    28857: gap of unknown length
*    28858    30242: contig of 1385 bp in length
*    30243    30342: gap of unknown length
*    30343    32113: contig of 1771 bp in length
*    32114    32213: gap of unknown length
*    32214    34175: contig of 1962 bp in length
*    34176    34275: gap of unknown length
*    34276    35470: contig of 1195 bp in length
*    35471    35570: gap of unknown length
*    35571    37322: contig of 1752 bp in length
*    37323    37422: gap of unknown length
*    37423    38791: contig of 1369 bp in length
*    38792    38891: gap of unknown length
*    38892    40361: contig of 1470 bp in length
*    40362    40461: gap of unknown length
*    40462    41995: contig of 1534 bp in length
*    41996    42095: gap of unknown length
*    42096    44157: contig of 2062 bp in length
*    44158    44257: gap of unknown length
*    44258    45517: contig of 1260 bp in length
*    45518    45617: gap of unknown length
*    45618    46970: contig of 1353 bp in length
*    46971    47070: gap of unknown length
*    47071    49593: contig of 2523 bp in length
*    49594    49693: gap of unknown length
*    49694    51492: contig of 1799 bp in length
*    51493    51592: gap of unknown length
*    51593    53277: contig of 1685 bp in length
*    53278    53377: gap of unknown length
*    53378    54782: contig of 1405 bp in length
*    54783    54882: gap of unknown length
*    54883    57021: contig of 2139 bp in length
*    57022    57121: gap of unknown length
*    57122    59001: contig of 1880 bp in length
*    59002    59101: gap of unknown length
*    59102    60121: contig of 1020 bp in length
*    60122    60221: gap of unknown length
*    60222    62559: contig of 2338 bp in length
*    62560    62659: gap of unknown length
*    62660    64766: contig of 2107 bp in length
*    64767    64866: gap of unknown length
*    64867    66686: contig of 1820 bp in length
*    66687    66786: gap of unknown length
*    66787    68406: contig of 1620 bp in length
*    68407    68506: gap of unknown length
*    68507    70230: contig of 1724 bp in length
*    70231    70330: gap of unknown length
*    70331    72687: contig of 2357 bp in length
*    72688    72787: gap of unknown length
*    72788    73961: contig of 1174 bp in length
*    73962    74061: gap of unknown length
*    74062    75594: contig of 1533 bp in length
*    75595    75694: gap of unknown length
*    75695    77376: contig of 1682 bp in length
*    77377    77476: gap of unknown length
*    77477    79269: contig of 1793 bp in length
*    79270    79369: gap of unknown length
*    79370    81510: contig of 2141 bp in length
*    81511    81610: gap of unknown length
*    81611    83780: contig of 2170 bp in length
*    83781    83880: gap of unknown length
*    83881    86085: contig of 2205 bp in length
*    86086    86185: gap of unknown length
*    86186    88050: contig of 1865 bp in length
*    88051    88150: gap of unknown length
*    88151    91051: contig of 2901 bp in length
*    91052    91151: gap of unknown length
*    91152    93335: contig of 2184 bp in length
*    93336    93435: gap of unknown length
*    93436    96781: contig of 3346 bp in length

  Query Match           16.0%;  Score 38.4;  DB 2;  Length 192153;
  Best Local Similarity   55.1%;  Pred. No. 3.3;
  Matches   75;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy      101 TTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAACAAAAT 160
            ||||||   ||  || ||  |  | |  |          || |||||||| ||    || |
Db    31052 TTTTTATATACTTCTTTATACTGAATATTCTACTGCTACTGAAAATACAACTGGAAATCT 30993

Qy      161 GATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAACTAATAAACGCTCTCGCACC 220
            || |  | ||| || | ||  | | | | ||||   ||  | | || | | | | | |||
Db    30992 TATAGTGAGTGAGTGCTTTGGTTTAATTTTGAAACACACACACACACACACACACA 30933

Qy      221 CGCAATTATATTTTGT 236
            | |||||| | | | |
Db    30932 CACAATTACAGTGTTT 30917


RESULT 11
AF022972/c
LOCUS       AF022972               32940 bp    DNA     linear   INV 17-MAY-2002
DEFINITION  Caenorhabditis elegans cosmid F10D2, complete sequence.
ACCESSION   AF022972
VERSION     AF022972.1  GI:2384818
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 32940)
  AUTHORS   Waterston,R.
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
  PUBMED    9851916
REFERENCE   2  (bases 1 to 32940)
  AUTHORS   Graves,T., Wohldmann,P. and Gillam,B.
  TITLE     The sequence of C. elegans cosmid F10D2
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 32940)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
COMMENT     Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University
                  St. Louis , MO  63110, USA, and
                  Sanger Centre, Hinxton Hall
                  Cambridge CB10 1RQ, England
                  email: rw@nematode.wustl.edu and jes@sanger.ac.uk

            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.
```

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

For a graphical representation of this cosmid sequence and its analysis see:
http://www.wormbase.org/db/seq/sequence?name=F10D2;class=Sequence

NEIGHBORING COSMID INFORMATION

The 5' cosmid is T01C4, 997 bp overlap; the 3' cosmid is F09G2, 500 bp overlap. Actual start of this cosmid is at base position 798 of F10D2; actual end is at 6797 of F09G2.

NOTES:

Coding seqences below are the result of integration and manual review of the following data : computer analysis using the program Genefinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yuji Kohara (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C. elegans ORFeome cloning project (http://worfdb.dfci.harvard.edu/), similarity to other proteins from BlastX analyses (http://blast.wustl.edu/), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

```
FEATURES             Location/Qualifiers
     source          1. .32940
                     /organism="Caenorhabditis elegans"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /chromosome="V"
                     /clone="F10D2"
     gene            494. .3628
                     /gene="F10D2.3"
                     /note="for a graphical representation of this gene see:
                     http://www.wormbase.org/db/seq/sequence?name=F10D2.3;class
                     =Sequence"
     CDS             join(494. .603,660. .817,1066. .1134,1233. .1444,1558. .2076,
                     3092. .3163,3235. .3377,3424. .3628)
                     /gene="F10D2.3"
                     /codon_start=1
                     /product="Hypothetical protein F10D2.3"
                     /protein_id="AAC48234.1"
                     /db_xref="GI:2384821"
                     /translation="MLLIHYVNLTHFKTSHNEFWLDLKVFLAIECVLYFIDLFNMIFN
                     FVFLIRAHQFHFNFGCIYAWLFVINFTDNIALLLMKNIMIWEYVDDSNPTSNWLFFIT
                     MNITIACTFAAMCTLFFCAVERCFATVYICDYEKKQRKFISIMLNSMLTIFGFAVCGV
                     VTDKKNTVYLIIFLLVIDGIALMLHCFLQWLNKRIYVRLHDNVYITSYSLAQRFQVAE
                     NIVPTGKHHILISQKPHYEFQMMNNVIYYMGFMNMTVVLSVLFSSFDLTPEHNFVITL
                     CLEVSIFIYSFCYPVIMYHSCDRWKTEINLFLKFVKIRNPVINSKVYPIRNSYGISMS
                     EENTMSNHFDHLRDAWETNQGKGKFQYRTHHSKQKKEVSIEKFLKMIRFHNNAEPSEI
                     AWSCMSPISRCCGIGCCPKTSKSQSESIFHSVFGYMCGGFMFLCCLVLLLSLVMFLCN
                     DVWHRYIPWFPSSPRPTATPRSETHEMQELNPSTHRSPSRLHLAV"
     gene            complement(4273. .6363)
                     /gene="str-112"
                     /note="for a graphical representation of this gene see:
                     http://www.wormbase.org/db/seq/sequence?name=F10D2.4;class
                     =Sequence"
     CDS             complement(join(4273. .4373,4427. .4604,4712. .4837,
                     4883. .4957,5004. .5070,5123. .5257,5322. .5485,6169. .6363))
                     /gene="str-112"
                     /codon_start=1
                     /product="C. elegans STR-112 protein (corresponding
                     sequence F10D2.4)"
                     /protein_id="AAC48242.1"
                     /db_xref="GI:2384829"
                     /translation="MSGQLWLALVDAADMVGFTLTISINIILLGLIRTRGKTLGTYKY
                     LMSFFSFFSIFYAIVESILRPIMHIENTTFFLISRKRFDYSTRLGKINSAFYCACFAT
                     SFVLSAVHFVYRYFAACKPNLLRLFNLPHLLLWPLMCSIPVTAWASVSYFLYPDTEYT
                     EAAVTYVLKTHYEVIKKENVSYIAYVYYQYENGERHIYIKNLLGCFVHYFVMSMTFVV
                     VFYCGFSTWWTIREHRGASDRTRHLHRQLFKALVFQTLVPSIFMYIPTGVMFIAPFFD
                     INLNANANFIVFCSFLYPGLDPLILIFIIREFRVTILNIIRGNERGNAVGEAYSTSRI
                     KSSQPAAVNLSG"
     gene            complement(7548. .9888)
                     /gene="F10D2.5"
                     /note="for a graphical representation of this gene see:
                     http://www.wormbase.org/db/seq/sequence?name=F10D2.5;class
                     =Sequence"
     CDS             complement(join(7548. .7722,7838. .7933,8029. .8165,
                     8211. .8296,8419. .8606,8652. .8887,8954. .9096,9143. .9253,
                     9387. .9609,9655. .9888))
                     /gene="F10D2.5"
                     /note="strong similarity to UDP-glucuronosyltransferases"
                     /codon_start=1
                     /product="Hypothetical protein F10D2.5"
                     /protein_id="AAC48235.1"
                     /db_xref="GI:2384822"
                     /translation="MKGFSLYLYYLLLISSTTSYNILVFCPLFGHSHSTFFGRLADIL
                     TEAGHNVTLFSPTIIDEFRNYSYTKLTKDVAYLDPSPELKAIGDLIAGNKRWWNQEFS
                     VFEIPQTTRFFKSITREQNNVLANNLALLDGLKQKKFDLILFESVFTCALPLMEYLEI
                     KSFAVAQSIAYEPITLGALGDTVMTSHIPNLMAQSSDKMSLCERMLNTLIQAMLNVYG
                     MTPEYYQSYTNPEKKIYTLEKLSEASFIFMNSNPFLDFPRATITKNIQIGGISVNLDT
                     LKSSGKLTEEWDQILNLREKTLLVSFGSVILSQDMPFAYKVGLTNAMKQLNDVTFIWK
                     YEGDDKKEFANGIKNIHFSKWVPQRELLGNNAFNFTEIHIYVTESDPRLSAFMTHGGL
                     GSVNEVSYFGKPTIMCPLSGDQMRNAKMLERHNGSIEFSKYDLHNEKVVANAFRKILY
                     DESYTLSAKRLSEHLQFQPVKPKELFLKHVEFAARFGKLPSLDPYSRQMGFIQYFLID
                     IAFIVLTVVLTIVSVLYFVLKFVASCLKAKIKTD"
     gene            complement(10003. .13250)
                     /gene="F10D2.6"
                     /note="for a graphical representation of this gene see:
                     http://www.wormbase.org/db/seq/sequence?name=F10D2.6;class
                     =Sequence"
     CDS             complement(join(10003. .10192,10938. .11033,11078. .11214,
                     11262. .11347,11394. .11533,11579. .11951,12253. .12363,
                     12414. .12723,13107. .13250))
                     /gene="F10D2.6"
                     /note="strong similarity to UDP-glucuronosyltransferases;
                     coded for by the following C. elegans cDNAs: yk1172e10.3,
                     yk1172e10.5"
                     /codon_start=1
                     /product="Hypothetical protein F10D2.6"
                     /protein_id="AAC48236.2"
                     /db_xref="GI:20901939"
                     /translation="MKILILLFACFNVNFSYNFLIFSPLFGHSHTTFFAKIADTLSEA
                     GHNVTFFTPTIIRKFSKINYVKHTKHVIHLEPSEKLERYGNQMEDVDISRFWTDDSSM
                     AEMFPMIKLFNEMFAEQAFVLGQNLDVLDELKEMKFDVMIFERFAECAYPLLEYLEIE
                     TFIPSSSLLFDPNLLISIGEPRIPSTVPLPMSKFTDKMTMTERLLNAIAVPIFHSLLP
                     THEFKSFRPPYQAIDIQKMEHLSSFILSNSNPYIDYPRPTLEKNVQIGGISVDVENLK
                     IHKVNEEWDRILDMRQKTILVSFGSVMLSKDMPVENKKILAKTMKQFPEVTFIWKYEF
                     NDTDLFASETENIHFSSWVPQTALLADRRLTAFFTHAGLGSVNEVSYLGKPSIMCPIF
                     ADQMRNAKMLARHNGSIEISKYDLGNGEKIEKTLRTILFDDSYRLSAEKLAHQLANQP
                     VKPKELLLRHAEFAAQFGRLPSLDPYSRQLSFVEYFLIDVAAILIFSSLLIVVVLFQI
                     IKRLINFLPVTVDSRKVKTN"
     gene            complement(13602. .15416)
                     /gene="F10D2.7"
                     /note="for a graphical representation of this gene see:
                     http://www.wormbase.org/db/seq/sequence?name=F10D2.7;class
                     =Sequence"
     CDS             complement(join(13602. .13776,13825. .14057,14101. .14186,
                     14250. .14873,14917. .15226,15273. .15416))
                     /gene="F10D2.7"
                     /note="strong similarity to UDP-glucuronosyltransferases"
                     /codon_start=1
                     /product="Hypothetical protein F10D2.7"
                     /protein_id="AAC48237.1"
                     /db_xref="GI:2384824"
                     /translation="MLLLALLFTLYANVSSFNFLVFCPLFAHSHSKFFATIADSLTDA
                     GHNVAFFTPIIVEKYRNFNYTKSTKDVVYMEPSKKLKEYGRQMSTGDFVRYWTEDSTA
                     TEIITVVRIFQKMYNEQGTVMKDNIEILNDLKKRKFDVIIFEAFIFSAYPLMDYLEIP
```

```
                     TFIPSLSVAHDTALAVEIGEPVMPSVVSDCLSPFGDKLSITERAFNAFMVPYFDFMIG
                     YPTHRSFKPPYNPIDIRSIVPEASFVFLNSNPFVDFPRPTLTKTIQIGGISVDSNELR
                     SQKLDETWSEILKLREKTMLVSFGSMLYSKDMPFKNKIALKNAMEKFKNVTFIWKYED
                     DSSDEFAKGIENIHFAKWVPQTALLADSRLSAFLTHAGLGSITELSYLGKPAILCPQL
                     FDQMRNTKMLVRHNGSIELSKYDLGKSEKIIEAFQAILFDSSYAKNAQKLAEQLENQP
                     IKPKEMMVKHAEYAARFGRLPSLDPYSRKMSFVEYFLLDVAAICVFILTFVVFVVTKL
                     VQSLFDIFLIKRKTE"
     gene            complement(16634..18035)
                     /gene="F10D2.12"
                     /note="for a graphical representation of this gene see:
                     http://www.wormbase.org/db/seq/sequence?name=F10D2.12;clas
                     s=Sequence"
     CDS             complement(join(16634..16808,16852..17084,17132..17217,
                     17417..18035))
                     /gene="F10D2.12"

  Query Match             15.8%;  Score 38;  DB 3;  Length 32940;
  Best Local Similarity   55.2%;  Pred. No. 4.7;
  Matches   74;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy         77 CATACTCCGAAACCACTGGACGGATTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCT 136
              || |  ||||      ||  |  ||||  | | |||||||  ||| |    ||||  |||||
Db       9152 CACATACCGAGTGAGCTATAATGATTAATTTTATGTTTGACTTAAAAATTTGAAGACCT 9093

Qy        137 GCCTAAAAATACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAACA 196
                 |      | |  || |  | |||||||   | ||||| ||  ||       | | | |  ||
Db       9092 GATGGCACAATCTTCTGATAAAATGAGCCTATGCGAAAGAATGCTGAACACTCTGATTCA 9033

Qy        197 AGAACTAATAAACG 210
               || | |  | || |
Db       9032 AGCAATGCTCAATG 9019


RESULT 12
AC111374/c
LOCUS       AC111374              114473 bp    DNA     linear   HTG 13-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-81E15, *** SEQUENCING IN PROGRESS
            ***, 59 unordered pieces.
ACCESSION   AC111374
VERSION     AC111374.2  GI:21735902
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 114473)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 114473)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 114473)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 12, 2002 this sequence version replaced gi:18701138.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GMDD
                Center clone name: CH230-81E15
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 69854 bases at least Q40
                Consensus quality: 74696 bases at least Q30
                Consensus quality: 78828 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *    (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 59 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      1324: contig of 1324 bp in length
            *     1325      1424: gap of unknown length
            *     1425      2908: contig of 1484 bp in length
            *     2909      3008: gap of unknown length
            *     3009      4062: contig of 1054 bp in length
            *     4063      4162: gap of unknown length
            *     4163      5249: contig of 1087 bp in length
            *     5250      5349: gap of unknown length
            *     5350      6638: contig of 1289 bp in length
            *     6639      6738: gap of unknown length
            *     6739      7791: contig of 1053 bp in length
            *     7792      7891: gap of unknown length
            *     7892      8911: contig of 1020 bp in length
            *     8912      9011: gap of unknown length
            *     9012     10059: contig of 1048 bp in length
            *    10060     10159: gap of unknown length
            *    10160     11221: contig of 1062 bp in length
            *    11222     11321: gap of unknown length
```

```
*    11322    12458: contig of 1137 bp in length
*    12459    12558: gap of unknown length
*    12559    13663: contig of 1105 bp in length
*    13664    13763: gap of unknown length
*    13764    14901: contig of 1138 bp in length
*    14902    15001: gap of unknown length
*    15002    16670: contig of 1669 bp in length
*    16671    16770: gap of unknown length
*    16771    18314: contig of 1544 bp in length
*    18315    18414: gap of unknown length
*    18415    20193: contig of 1779 bp in length
*    20194    20293: gap of unknown length
*    20294    21742: contig of 1449 bp in length
*    21743    21842: gap of unknown length
*    21843    23705: contig of 1863 bp in length
*    23706    23805: gap of unknown length
*    23806    25889: contig of 2084 bp in length
*    25890    25989: gap of unknown length
*    25990    27690: contig of 1701 bp in length
*    27691    27790: gap of unknown length
*    27791    29133: contig of 1343 bp in length
*    29134    29233: gap of unknown length
*    29234    30464: contig of 1231 bp in length
*    30465    30564: gap of unknown length
*    30565    32347: contig of 1783 bp in length
*    32348    32447: gap of unknown length
*    32448    33461: contig of 1014 bp in length
*    33462    33561: gap of unknown length
*    33562    34870: contig of 1309 bp in length
*    34871    34970: gap of unknown length
*    34971    36226: contig of 1256 bp in length
*    36227    36326: gap of unknown length
*    36327    37738: contig of 1412 bp in length
*    37739    37838: gap of unknown length
*    37839    39756: contig of 1918 bp in length
*    39757    39856: gap of unknown length
*    39857    41572: contig of 1716 bp in length
*    41573    41672: gap of unknown length
*    41673    43234: contig of 1562 bp in length
*    43235    43334: gap of unknown length
*    43335    44957: contig of 1623 bp in length
*    44958    45057: gap of unknown length
*    45058    46349: contig of 1292 bp in length
*    46350    46449: gap of unknown length
*    46450    48713: contig of 2264 bp in length
*    48714    48813: gap of unknown length
*    48814    49911: contig of 1098 bp in length
*    49912    50011: gap of unknown length
*    50012    52028: contig of 2017 bp in length
*    52029    52128: gap of unknown length
*    52129    54226: contig of 2098 bp in length
*    54227    54326: gap of unknown length
*    54327    56358: contig of 2032 bp in length
*    56359    56458: gap of unknown length
*    56459    57923: contig of 1465 bp in length
*    57924    58023: gap of unknown length
*    58024    60157: contig of 2134 bp in length
*    60158    60257: gap of unknown length
*    60258    61265: contig of 1008 bp in length
*    61266    61365: gap of unknown length
*    61366    62546: contig of 1181 bp in length
*    62547    62646: gap of unknown length
*    62647    64341: contig of 1695 bp in length
*    64342    64441: gap of unknown length
*    64442    66383: contig of 1942 bp in length
*    66384    66483: gap of unknown length
*    66484    67902: contig of 1419 bp in length
*    67903    68002: gap of unknown length
*    68003    69421: contig of 1419 bp in length
*    69422    69521: gap of unknown length
*    69522    71916: contig of 2395 bp in length
*    71917    72016: gap of unknown length
*    72017    74548: contig of 2532 bp in length
*    74549    74648: gap of unknown length
*    74649    77103: contig of 2455 bp in length
*    77104    77203: gap of unknown length
*    77204    79285: contig of 2082 bp in length
*    79286    79385: gap of unknown length
*    79386    81755: contig of 2370 bp in length
*    81756    81855: gap of unknown length
*    81856    85438: contig of 3583 bp in length
*    85439    85538: gap of unknown length
*    85539    88732: contig of 3194 bp in length
*    88733    88832: gap of unknown length
*    88833    90592: contig of 1760 bp in length
*    90593    90692: gap of unknown length
*    90693    92908: contig of 2216 bp in length
*    92909    93008: gap of unknown length
*    93009    95062: contig of 2054 bp in length

  Query Match             15.7%;  Score 37.6;  DB 2;  Length 114473;
  Best Local Similarity   56.5%;  Pred. No. 5.6;
  Matches   70;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      102 TTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAACAAAATG 161
            |||| || | ||||||||     |  ||| | |    || |||    ||    | ||
Db    26389 TTTTTTGCTTTTTGTTTTGTTAAGTGGGGTTAGCAATGTAGAAAGGACACTGGAAGGATC 26330

Qy      162 ATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAACTAATAAACGCTCTCGCACCC 221
            ||| |   || | ||  |||||| || | ||  || ||     |||| || || |
Db    26329 ATTGTAGGCAAAAGAGATGGCGTTAGTAAAGAAAACAAAAGTCAAACTATCACGGAAGT 26270

Qy      222 GCAA 225
            ||||
Db    26269 GCAA 26266


RESULT 13
AC007514/c
LOCUS       AC007514               177720 bp    DNA     linear   PRI 26-AUG-1999
DEFINITION  Homo sapiens chromosome 14 clone BACs 209A20 and 396N8 map 14q31,
            complete sequence.
ACCESSION   AC007514
VERSION     AC007514.5  GI:5776573
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 177720)
  AUTHORS   Dickhoff,R., Madan,A., Qin,S., Abbasi,N., Baradarani,L.,
            Birditt,B., Bloom,S., Dors,M., Rowen,L., Harrison,G., James,R.,
            Lasky,S., Madan,A., Ratcliffe,A., Shaffer,T. and Hood,L.
  TITLE     Sequencing of human chromosome 14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 177720)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Dors,M., Dickhoff,R.,
            James,R., Lorentz,P., Lasky,S., Madan,A., Prescott,S., Ratcliffe,A.,
            Shaffer,T. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAY-1999) Multimegabase Sequencing Center, University
            of Washington, PO BOX 357730, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 177720)
  AUTHORS   Dickhoff,R., Madan,A., Qin,S., Abbasi,N., Baradarani,L.,
            Birditt,B., Bloom,S., Dors,M., Rowen,L., Harrison,G., James,R.,
            Lasky,S., Madan,A., Ratcliffe,A., Shaffer,T. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-AUG-1999) Multimegabase Sequencing Center, University
            of Washington, PO BOX 357730, Seattle, WA 98195, USA
COMMENT     On Aug 26, 1999 this sequence version replaced gi:5668760.
            This sequence is derived from the two haplotypes of RPCI-11.  No
            genes were found in this sequence.
FEATURES             Location/Qualifiers
     source          1..177720
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
                     /chromosome="14"
                     /map="14q31"
                     /clone="BACs 209A20 and 396N8"
                     /clone_lib="RPCI-11"
                     /note="These clones overlap BAC 526N18, Accession AF123462
                     and BAC 115A24, Accession AC007372"
     source          1..142843
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="BAC 209A20"
     source          10153..177720
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="BAC 396N8"
     misc_feature    1..8135
                     /note="overlap with BAC 526N18, Accession AF123462"
     variation       937
                     /note="526N18: g; 209A20: t"
                     /replace="t"
     variation       1703
                     /standard_name="t"
                     /note="526N18: a; 209A20: t"
     variation       1808
                     /note="526N18: c; 209A20: t"
                     /replace="t"
     variation       3576
                     /note="526N18: c; 209A20: t"
                     /replace="t"
     variation       7277
                     /note="526N18: t; 209A20: c"
                     /replace="c"
     variation       12486
                     /note="396N8: c: 209A20: a"
                     /replace="a"
     variation       13272
                     /note="396N8: a; 209A20: t"
                     /replace="t"
     variation       13489
                     /note="396N8: c; 209A20: t"
                     /replace="t"
     variation       13740
                     /note="396N8: a; 209A20: g"
                     /replace="g"
     variation       16546
                     /note="396N18: t; 209A20: c"
                     /replace="c"
     variation       17879
                     /note="396N8: c; 209A20: t"
                     /replace="t"
     variation       18411
                     /note="396N8: c; 209A20: t"
                     /replace="t"
     variation       19843
                     /note="396N8: a; 209A20: g"
                     /replace="g"
     variation       19882
                     /note="396N8: a; 209A20: g"
                     /replace="g"
     variation       20021
                     /note="396N8: g; 209A20: t"
                     /replace="t"
     variation       20398
                     /note="396N8: t; 209A20: c"
                     /replace="c"
     variation       20463
                     /note="396N8: a; 209A20: g"
                     /replace="g"
     variation       21631
                     /note="396N8: c; 209A20: t"
                     /replace="t"
     variation       21817
                     /note="396N8: c; 209A20: t"
                     /replace="t"
     variation       21968
                     /note="396N8: g; 209A20: t"
                     /replace="t"
     variation       22173
                     /note="396N8: g; 209A20: a"
                     /replace="a"
     variation       23292
                     /note="396N8: a; 209A20: t"
                     /replace="t"
     variation       24192
                     /note="396N8: a; 209A20: t"
                     /replace="t"
     variation       25565
                     /note="396N8: a; 209A20: g"
                     /replace="g"
     variation       26539
                     /note="396N8: t; 209A20: c"
                     /replace="c"
     variation       27327
                     /note="396N8: a; 209A20: g"
                     /replace="g"
     variation       28244
                     /note="396N8: a; 209A20: g"
                     /replace="g"
     variation       28528
                     /note="396N8: c; 209A20: t"
                     /replace="t"
     variation       28562
                     /note="396N8: g; 209A20: a"
                     /replace="a"
     variation       28998
                     /note="396N8: g; 209A20: a"
                     /replace="a"
     variation       29128
                     /note="396N8: g; 209A20: a"
                     /replace="a"
     variation       29373
                     /note="396N8: c; 209A20: a"
                     /replace="a"
     variation       29502
                     /note="396N8: a; 209A20: t"
                     /replace="t"
     variation       31890
                     /note="396N8: t; 209A20: c"
                     /replace="c"
     variation       34931
                     /note="396N18: c; 209A20: t"
                     /replace="t"
     variation       37453
                     /note="396N8: t; 209A20: c"
                     /replace="c"
     variation       37961
                     /note="396N8: g; 209A20: a"
                     /replace="a"
     variation       38392
                     /note="396N8: g; 209A20: a"
                     /replace="a"
     variation       38737
                     /note="396N8: c; 209A20: t"
                     /replace="t"
     variation       39027
                     /note="396N8: a; 209A20: g"
                     /replace="g"
     variation       40709
                     /note="396N8: a; 209A20: t"
                     /replace="t"
     variation       41384
                     /note="396N8: c; 209A20: a"
                     /replace="a"
     variation       41735
                     /note="396N18: a; 209A20: acaca"
                     /replace="acaca"
     variation       42385
```

```
                    /note="396N8: c; 209A20: t"
                    /replace="t"
     variation      44238
                    /note="396N8: g; 209A20: t"
                    /replace="t"
     variation      49415
                    /note="396N8: t; 209A20: c"
                    /replace="c"
     variation      49659
                    /note="396N8: c; 209A20: a"
                    /replace="a"
     variation      50923
                    /note="396N8: c; 209A20: g RFLP: HindIII"
                    /replace="g"
     variation      58154. .58160
                    /note="396N8: aaaaagg; 209A20: g"
                    /replace="g"

  Query Match             15.7%;  Score 37.6;  DB 9;  Length 177720;
  Best Local Similarity   59.3%;  Pred. No. 5.5;
  Matches   64;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy        98 GGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAACAA 157
                || || || |  | ||| ||||| ||   ||||  ||||   |   |||   || | | ||
Db      5922 GGCTTCTTGTTTTTTTTCTTTTACAGAAATTGGAAACCTTGTTCTAAAATTTATACAGAA 5863

Qy       158 AATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAACTAAT 205
             |   |  |  | ||| || || |     | || | ||| |||||| |||
Db      5862 AGGCAGGAAATCCGAAAGAGTAGAAAATTATTTAAAAAAGAACAAAT 5815


RESULT 14
AC125626/c
LOCUS       AC125626              190481 bp    DNA     linear   HTG 10-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-10D16, *** SEQUENCING IN PROGRESS
            ***, 80 unordered pieces.
ACCESSION   AC125626
VERSION     AC125626.2  GI:21630298
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 190481)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 190481)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 190481)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jun 30, 2002 this sequence version replaced gi:21629089.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GDNP
                Center clone name: CH230-10D16
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 109625 bases at least Q40
                Consensus quality: 120858 bases at least Q30
                Consensus quality: 126621 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 80 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *          1      1279: contig of 1279 bp in length
            *       1280      1379: gap of unknown length
            *       1380      2880: contig of 1501 bp in length
            *       2881      2980: gap of unknown length
            *       2981      4516: contig of 1536 bp in length
            *       4517      4616: gap of unknown length
            *       4617      6499: contig of 1883 bp in length
            *       6500      6599: gap of unknown length
            *       6600      8135: contig of 1536 bp in length
            *       8136      8235: gap of unknown length
            *       8236      9387: contig of 1152 bp in length
            *       9388      9487: gap of unknown length
            *       9488     10819: contig of 1332 bp in length
            *      10820     10919: gap of unknown length
            *      10920     12013: contig of 1094 bp in length
            *      12014     12113: gap of unknown length
            *      12114     13223: contig of 1110 bp in length
            *      13224     13323: gap of unknown length
            *      13324     14891: contig of 1568 bp in length
            *      14892     14991: gap of unknown length
```

```
*    14992    16206: contig of 1215 bp in length
*    16207    16306: gap of unknown length
*    16307    17471: contig of 1165 bp in length
*    17472    17571: gap of unknown length
*    17572    18602: contig of 1031 bp in length
*    18603    18702: gap of unknown length
*    18703    20436: contig of 1734 bp in length
*    20437    20536: gap of unknown length
*    20537    21612: contig of 1076 bp in length
*    21613    21712: gap of unknown length
*    21713    23092: contig of 1380 bp in length
*    23093    23192: gap of unknown length
*    23193    24202: contig of 1010 bp in length
*    24203    24302: gap of unknown length
*    24303    25451: contig of 1149 bp in length
*    25452    25551: gap of unknown length
*    25552    26958: contig of 1407 bp in length
*    26959    27058: gap of unknown length
*    27059    29049: contig of 1991 bp in length
*    29050    29149: gap of unknown length
*    29150    30242: contig of 1093 bp in length
*    30243    30342: gap of unknown length
*    30343    31744: contig of 1402 bp in length
*    31745    31844: gap of unknown length
*    31845    32906: contig of 1062 bp in length
*    32907    33006: gap of unknown length
*    33007    34764: contig of 1758 bp in length
*    34765    34864: gap of unknown length
*    34865    36081: contig of 1217 bp in length
*    36082    36181: gap of unknown length
*    36182    37974: contig of 1793 bp in length
*    37975    38074: gap of unknown length
*    38075    39488: contig of 1414 bp in length
*    39489    39588: gap of unknown length
*    39589    40724: contig of 1136 bp in length
*    40725    40824: gap of unknown length
*    40825    42022: contig of 1198 bp in length
*    42023    42122: gap of unknown length
*    42123    44712: contig of 2590 bp in length
*    44713    44812: gap of unknown length
*    44813    46747: contig of 1935 bp in length
*    46748    46847: gap of unknown length
*    46848    48457: contig of 1610 bp in length
*    48458    48557: gap of unknown length
*    48558    50410: contig of 1853 bp in length
*    50411    50510: gap of unknown length
*    50511    51683: contig of 1173 bp in length
*    51684    51783: gap of unknown length
*    51784    53079: contig of 1296 bp in length
*    53080    53179: gap of unknown length
*    53180    54864: contig of 1685 bp in length
*    54865    54964: gap of unknown length
*    54965    56334: contig of 1370 bp in length
*    56335    56434: gap of unknown length
*    56435    58299: contig of 1865 bp in length
*    58300    58399: gap of unknown length
*    58400    60583: contig of 2184 bp in length
*    60584    60683: gap of unknown length
*    60684    63235: contig of 2552 bp in length
*    63236    63335: gap of unknown length
*    63336    64948: contig of 1613 bp in length
*    64949    65048: gap of unknown length
*    65049    66722: contig of 1674 bp in length
*    66723    66822: gap of unknown length
*    66823    68638: contig of 1816 bp in length
*    68639    68738: gap of unknown length
*    68739    70615: contig of 1877 bp in length
*    70616    70715: gap of unknown length
*    70716    72402: contig of 1687 bp in length
*    72403    72502: gap of unknown length
*    72503    74170: contig of 1668 bp in length
*    74171    74270: gap of unknown length
*    74271    77608: contig of 3338 bp in length

*    77609    77708: gap of unknown length
*    77709    80029: contig of 2321 bp in length
*    80030    80129: gap of unknown length
*    80130    82227: contig of 2098 bp in length
*    82228    82327: gap of unknown length
*    82328    84026: contig of 1699 bp in length
*    84027    84126: gap of unknown length
*    84127    87363: contig of 3237 bp in length
*    87364    87463: gap of unknown length
*    87464    89262: contig of 1799 bp in length
*    89263    89362: gap of unknown length
*    89363    90768: contig of 1406 bp in length
*    90769    90868: gap of unknown length
*    90869    92638: contig of 1770 bp in length

  Query Match              15.7%;  Score 37.6;  DB 2;  Length 190481;
  Best Local Similarity    56.5%;  Pred. No. 5.5;
  Matches    70;  Conservative    0;  Mismatches    54;  Indels    0;  Gaps    0;

Qy      102 TTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAACAAAATG 161
            |||| || | |||||||          | ||| | |   || |||   ||    | ||
Db   178820 TTTTTTGCTTTTTGTTTTGTTAAGTGGGGTTAGCAATGTAGAAAGGACACTGGAAGGATC 178761

Qy      162 ATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAACTAATAAACGCTCTCGCACCC 221
            ||| |  || | ||  |||||| || | || ||  |||| || | |
Db   178760 ATTGTAGGCAAAAGAGATGGCGTTAGTAAAGAAAACAAAAGTCAAACTATCACGGAAGT 178701

Qy      222 GCAA 225
            ||||
Db   178700 GCAA 178697


RESULT 15
AC094816
LOCUS       AC094816               220181 bp    DNA     linear   HTG 08-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-5J4, *** SEQUENCING IN PROGRESS ***,
            83 unordered pieces.
ACCESSION   AC094816
VERSION     AC094816.4  GI:21702754
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 220181)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.P., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwite,K.J., Draper,H., Dugan-Rocha,S., Dumpati,K.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,P., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Lu,J., Lu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., Mcleod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
```

```
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shoobtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,R., Stone,R.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warred,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wlecnyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,T., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 220181)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 220181)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (08-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 7, 2002 this sequence version replaced gi:21700357.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GBLG
                Center clone name: CH230-524
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap, version  0.990329
                Consensus quality: 144990 bases at least Q40
                Consensus quality: 152139 bases at least Q30
                Consensus quality: 157114 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 83 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1012: contig of 1012 bp in length
            *    1013     1112: gap of unknown length
            *    1113     2218: contig of 1106 bp in length
            *    2219     2318: gap of unknown length
            *    2319     3478: contig of 1160 bp in length
            *    3479     3578: gap of unknown length
            *    3579     4630: contig of 1052 bp in length
            *    4631     4730: gap of unknown length
            *    4731     5825: contig of 1095 bp in length
            *    5826     5925: gap of unknown length
            *    5926     6995: contig of 1070 bp in length
            *    6996     7095: gap of unknown length
            *    7096     8362: contig of 1267 bp in length
            *    8363     8462: gap of unknown length
            *    8463     9619: contig of 1157 bp in length
            *    9620     9719: gap of unknown length
            *    9720    10755: contig of 1036 bp in length
            *   10756    10855: gap of unknown length
            *   10856    11870: contig of 1015 bp in length
            *   11871    11970: gap of unknown length
            *   11971    13282: contig of 1312 bp in length
            *   13283    13382: gap of unknown length
            *   13383    14394: contig of 1012 bp in length
            *   14395    14494: gap of unknown length
            *   14495    15562: contig of 1068 bp in length
            *   15563    15662: gap of unknown length
            *   15663    16984: contig of 1322 bp in length
            *   16985    17084: gap of unknown length
            *   17085    18247: contig of 1163 bp in length
            *   18248    18347: gap of unknown length
            *   18348    19428: contig of 1081 bp in length
            *   19429    19528: gap of unknown length
            *   19529    20593: contig of 1065 bp in length
            *   20594    20693: gap of unknown length
            *   20694    21937: contig of 1244 bp in length
            *   21938    22037: gap of unknown length
            *   22038    23353: contig of 1316 bp in length
            *   23354    23453: gap of unknown length
            *   23454    24950: contig of 1497 bp in length
            *   24951    25050: gap of unknown length
            *   25051    26476: contig of 1426 bp in length
            *   26477    26576: gap of unknown length
            *   26577    27609: contig of 1033 bp in length
            *   27610    27709: gap of unknown length
            *   27710    29121: contig of 1412 bp in length
            *   29122    29221: gap of unknown length
            *   29222    30257: contig of 1036 bp in length
            *   30258    30357: gap of unknown length
            *   30358    31806: contig of 1449 bp in length
            *   31807    31906: gap of unknown length
            *   31907    32970: contig of 1064 bp in length
            *   32971    33070: gap of unknown length
            *   33071    34431: contig of 1361 bp in length
            *   34432    34531: gap of unknown length
            *   34532    36131: contig of 1600 bp in length
            *   36132    36231: gap of unknown length
            *   36232    37546: contig of 1315 bp in length
            *   37547    37646: gap of unknown length
            *   37647    39096: contig of 1450 bp in length
            *   39097    39196: gap of unknown length
            *   39197    41216: contig of 2020 bp in length
            *   41217    41316: gap of unknown length
            *   41317    42842: contig of 1526 bp in length
            *   42843    42942: gap of unknown length
            *   42943    45188: contig of 2246 bp in length
            *   45189    45288: gap of unknown length
            *   45289    46377: contig of 1089 bp in length
            *   46378    46477: gap of unknown length
            *   46478    47860: contig of 1383 bp in length
            *   47861    47960: gap of unknown length
            *   47961    50329: contig of 2369 bp in length
            *   50330    50429: gap of unknown length
            *   50430    52308: contig of 1879 bp in length
            *   52309    52408: gap of unknown length
            *   52409    53931: contig of 1523 bp in length
            *   53932    54031: gap of unknown length
            *   54032    55839: contig of 1808 bp in length
            *   55840    55939: gap of unknown length
            *   55940    57357: contig of 1418 bp in length
            *   57358    57457: gap of unknown length
            *   57458    59451: contig of 2034 bp in length
            *   59492    59591: gap of unknown length
            *   59592    61398: contig of 1807 bp in length
            *   61399    61498: gap of unknown length
            *   61499    63101: contig of 1603 bp in length
            *   63102    63201: gap of unknown length
            *   63202    66466: contig of 3265 bp in length
            *   66467    66566: gap of unknown length
            *   66567    68485: contig of 1919 bp in length
            *   68486    68585: gap of unknown length
            *   68586    69603: contig of 1018 bp in length
            *   69604    69703: gap of unknown length
```

```
            *    69704    72322: contig of 2619 bp in length
            *    72323    72422: gap of unknown length
            *    72423    74890: contig of 2468 bp in length
            *    74891    74990: gap of unknown length
            *    74991    77937: contig of 2947 bp in length
            *    77938    78037: gap of unknown length
            *    78038    80271: contig of 2234 bp in length
            *    80272    80371: gap of unknown length
            *    80372    83452: contig of 3081 bp in length
            *    83453    83552: gap of unknown length
            *    83553    84865: contig of 1313 bp in length
            *    84866    84965: gap of unknown length
            *    84966    87255: contig of 2290 bp in length
            *    87256    87355: gap of unknown length
            *    87356    89706: contig of 2351 bp in length

  Query Match             15.6%;  Score 37.4;  DB 2;  Length 220181;
  Best Local Similarity   48.5%;  Pred. No. 6.1;
  Matches   95;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy      28 AAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATACTCCGAA 87
           ||||  || |   ||    ||  ||  || || ||| |||| |              ||
Db    1295 AAAAAAATTTTCCGCAGGGGGGGGGGGGGGAAAAATATTTTTGGGGGGGGGNNNAAAA 1354

Qy      88 ACCACTGGACGGATTTTATGATGTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATA 147
           |  ||     |        |  | ||| |||| | |||         ||  ||  |||| |
Db    1355 AAAACCCCCCCCCCCCCCTTTTTTTTTTTTTTGTGGCCCCCCCCCCCCCCACAAAAAA 1414

Qy     148 CTACAAACAAAATGATTATTGCGAGAGCGTGGGCGTTAATATAAAACAAGAACTAATAA 207
           |  ||| |||| |   || | | | | ||| || | | |||| || ||  || ||
Db    1415 AAAAAAAAAAAAAAAATTTGGGGGGGGGGGGGGTTTAAAAAAAAAAAAAAAAAAAA 1474

Qy     208 ACGCTCTCGCACCCGC 223
           |  | | ||| |
Db    1475 NNTTTTCCCCCCCCC 1490


Search completed: May 12, 2003, 01:46:42
Job time : 1711 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 00:45:59 ; Search time 422.571 Seconds
(without alignments)
1279.025 Million cell updates/sec

Title: US-09-980-277-1
Perfect score: 240
Sequence: 1 gatctgaatccgaccaacg.....cgcaattatatttgttctt 240

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44.2	18.4	120	22	AAI70400	Schistosoma manson
2	44.2	18.4	6494	21	AAA46170	GFP-Sm1-7 fusion p
c 3	36.6	15.2	6134	22	AAS46564	Tumour suppressor
4	34.2	14.3	581	24	ABN60970	Human cancer relat
c 5	34	14.2	1396	23	ABL17579	Drosophila melanog
c 6	34	14.2	7729	23	ABL17578	Drosophila melanog
7	33.8	14.1	1020	16	AAQ75857	Haemolysin regulat
8	33.6	14.0	3363	20	AAX61476	B. burgdorferi ant
9	33.6	14.0	3441	20	AAX61475	B. burgdorferi ant
10	33.6	14.0	10723	22	AAD14603	Dengue virus (DEN)
11	33.6	14.0	10723	22	AAD14604	Dengue virus (DEN)
12	33.6	14.0	10735	22	AAD14601	Wild-type, virulen
13	33.6	14.0	10735	22	AAD14602	Attenuated, vaccin
c 14	33.4	13.9	1860	24	ABM70879	Streptococcus poly
c 15	33.4	13.9	1950	24	ABN66701	Streptococcus poly
16	33.2	13.8	10949	22	AAL07031	Human reproductive
c 17	33.2	13.8	10949	22	AAK79984	Human immune/haema
18	33.2	13.8	10949	22	AAI62709	Human breast or ov
19	33.2	13.8	10951	22	AAL07030	Human reproductive
c 20	33.2	13.8	10951	22	AAK79983	Human immune/haema
21	33.2	13.8	10951	22	AAI62708	Human breast or ov
22	33	13.8	39746	23	ABL11398	Drosophila melanog
c 23	32.4	13.5	8923	22	AAS46340	Tumour suppressor
24	32.4	13.5	48551	24	AAS20800	Clostridium diffic
25	32.2	13.4	964	24	ABQ70115	Listeria monocytog
26	32.2	13.4	964	24	ABQ71100	Listeria monocytog
c 27	32.2	13.4	2440	21	AAC59153	Human secreted pro
c 28	32.2	13.4	9359	20	AAX20262	Borrelia burgdorfe
29	32	13.3	83391	24	ABQ67094	Human angiogenesis
30	31.8	13.2	580073	18	AAT58840	Mycoplasma genital
31	31.6	13.2	621	23	ABV54682	Human prostate exp
32	31.6	13.2	669	24	ABN68671	Streptococcus poly
c 33	31.6	13.2	1933	22	AAK85566	Human immune/haema
c 34	31.6	13.2	2303	21	AAF18094	Lung cancer associ
c 35	31.6	13.2	10616	22	AAK85565	Human immune/haema
c 36	31.6	13.2	10616	22	AAK85567	Human immune/haema
c 37	31.6	13.2	40324	24	ABQ67149	Human angiogenesis
c 38	31.2	13.0	581	24	ABQ33362	Oligonucleotide fo
39	31.2	13.0	581	24	ABQ33363	Oligonucleotide fo
c 40	31.2	13.0	1888	22	AAR29873	C albicans apoptos
c 41	31.2	13.0	5611	21	AAA57948	5611 bp Candida al
42	31.2	13.0	12393	24	ABL33263	Human immune syste
43	31.2	13.0	640681	24	ABA92787	Buchnera sp. genom
c 44	30.8	12.8	1984	21	AAC76857	Human ORFX ORF2412
45	30.8	12.8	6931	22	ABA17762	Human nervous syst

ALIGNMENTS

RESULT 1
AAI70400
ID AAI70400 standard; DNA; 120 BP.
XX
AC AAI70400;
XX
DT 07-JAN-2002 (first entry)
XX
DE Schistosoma mansoni repeater unit DNA.
XX
KW Schistosomiasis; infection; diagnosis; ds.
XX
OS Schistosoma mansoni.
XX
FH Key Location/Qualifiers
FT primer_bind complement (1..19)
FT /*tag= a
FT /note= "primer of AAI70401"
FT primer_bind 91..110
FT /*tag= b
FT /note= "primer of AAI70402"
XX
PN WO200175148-A1.
XX
PD 11-OCT-2001.
XX
PF 04-APR-2001; 2001WO-BR00035.
XX
PR 04-APR-2000; 2000BR-0001536.
XX
PA (FIOC-) FIOCRUZ FUNDACAO CRUZ OSWALDO.
XX

```
PI   Teles Rabello AL,  Dias Neto E,  Pontes LA;
XX
DR   WPI; 2001-648561/74.
XX
PT   Diagnosing infection by parasites of the Schistosoma sp. in a
PT   biological sample especially useful in cases of low infection
PT   intensity, comprises detecting a specific region of the DNA of
PT   Schistosoma by polymerase chain reaction -
XX
PS   Claim 1; Page 22; 34pp; English.
XX
CC   The present sequence is that of a highly repeated short DNA
CC   sequence in the genome of Schistosoma mansoni.  PCR primers
CC   (see AAI70401 and AAI70402) that flank this sequence are used in
CC   the method of the invention for diagnosing Schistosoma infection.
CC   The method involves collection of the sample to be examined,
CC   extraction of Schistosoma sp. DNA, amplification by PCR, separation
CC   of PCR products by electrophoresis, and detection e.g. by colouring
CC   with silver salts.  A kit for diagnosing infection, which includes
CC   the primers, is claimed.  The method is useful for detecting
CC   Schistosoma sp. parasites by detecting the DNA of the parasite in a
CC   biological sample.  It is especially useful in cases of low
CC   infection intensity for which parasitological stool tests
CC   demonstrate little sensitivity.
XX
SQ   Sequence 120 BP; 36 A; 20 C; 26 G; 38 T; 0 other;

  Query Match           18.4%;  Score 44.2;  DB 22;  Length 120;
  Best Local Similarity   93.9%;  Pred. No. 0.00092;
  Matches   46;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      81 CTCCGAAACCACTGGACGGATTTTTATGATGTTGTTTTAGAGGCTTTG 129
           |||||||||||||||||||||||||||||||||||||||||||  ||||
Db      41 CTCCGAAACCACTGGACGGATTTTTATGATGTTGTTTTAGATTATTTG 89


RESULT 2
AAA46170
ID   AAA46170 standard; DNA; 6494 BP.
XX
AC   AAA46170;
XX
DT   27-SEP-2000  (first entry)
XX
DE   GFP-Sm1-7 fusion protein construct.
XX
KW   GFP-Sm1-7 fusion construct; circular; green fluorescent protein;
KW   glutathione S-transferase; eukaryotic diploid multicellular parasite;
KW   universal graft; transgenic eukaryotic parasite; acquired deficiency;
KW   genetic deficiency; hormone deficiency; metabolic deficiency;
KW   haematological deficiency; immunological deficiency; immunotherapy;
KW   anti-microbial therapy; anti-cancer therapy; drug addiction;
KW   poisoning condition; geriatric condition; ds.
XX
OS   Synthetic.
XX
PN   WO200032804-A1.
XX
PD   08-JUN-2000.
XX
PF   01-DEC-1999;   99WO-IL00651.
XX
PR   01-DEC-1998;   98US-0201850.
XX
PA   (YISS ) YISSUM RES & DEV CO.
XX
PI   Hamburger J,  Laban A;
XX
DR   WPI; 2000-412348/35.
XX
PT   Eukaryotic diploid multicellular parasite useful as universal grafts
PT   for in vivo delivery of beneficial gene products in humans and animals
PT   involves transformation with a transgene -
XX
PS   Example 2; Fig 6; 90pp; English.
XX
CC   This sequence represents a GFP-Sm1-7 fusion construct contained within a
CC   recombinant vector. This sequence contains the green fluorescent protein
CC   (GFP) coding sequence from Aequorea victoria, and the
CC   Sm1-7 coding sequence from Schistosoma, along with
CC   promoter sequences. The invention relates to a eukaryotic diploid
CC   multicellular parasite transformed with a transgene. Transgenic
CC   eukaryotic parasites are useful as universal grafts for in vivo delivery
CC   of beneficial gene product in humans and animals. The parasites can
CC   particularly be used for restoration of deficiencies whether acquired or
CC   genetic, such as hormone deficiencies, metabolic deficiencies,
CC   haematological deficiencies, immunological deficiencies, immunotherapy,
CC   anti-microbial therapy, anti-cancer therapy. They can also be used for
CC   treatment of drug addiction, of poisoning conditions, and for
CC   amelioration of geriatric conditions. Treatment of humans with in vivo
CC   transgenes are universally compatible, readily available and inexpensive.
CC   Genotypic alterations of the patients is avoided, reducing risks of
CC   mutagenesis and malignant transformation.
XX
SQ   Sequence 6494 BP; 1754 A; 1373 C; 1481 G; 1886 T; 0 other;

  Query Match           18.4%;  Score 44.2;  DB 21;  Length 6494;
  Best Local Similarity   93.9%;  Pred. No. 0.0027;
  Matches   46;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      81 CTCCGAAACCACTGGACGGATTTTTATGATGTTGTTTTAGAGGCTTTG 129
           |||||||||||||||||||||||||||||||||||||||||||  ||||
Db      41 CTCCGAAACCACTGGACGGATTTTTATGATGTTGTTTTAGATTATTTG 89


RESULT 3
AAS46564/C
ID   AAS46564 standard; DNA; 6134 BP.
XX
AC   AAS46564;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #286.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW   cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP02955.
XX
PR   15-MAR-2000; 2000DE-1013847.
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumour suppressor
PT   genes and oncogenes, useful in designing primers and probes for
PT   analysing diseases associated with cytosine methylation state e.g.
PT   cancer -
XX
PS   Claim 1; SEQ ID No 286; 27pp; English.
XX
```

```
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and
CC   oncogenes having a sequence taken from 536 (actually 533 since
CC   numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC   (Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC   peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC   form part of a set of probes for detecting the cytosine methylation state
CC   and/or single nucleotide polymorphisms and also to be used in an
CC   array for analysing diseases associated with CpG dinucleotides e.g.
CC   cancers and tumours. The probes can also be used in a method for
CC   ascertaining genetic and/or epigenetic parameters for the diagnosis
CC   and/or therapy of existing diseases or the predisposition to specific
CC   diseases, by analysing cytosine methylations. The parameters may be
CC   compared to another set of genetic and/or epigenetic parameters, the
CC   differences serving as basis for diagnosis and/or prognosis events which
CC   are disadvantageous to patients. The present sequence is one of the
CC   533 genomic sequences derived from tumour suppressor genes and
CC   oncogenes. Sequences with even numbered Seq ID numbers are the
CC   complementary sequence of the corresponding odd numbered sequence (e.g.
CC   ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
CC   is missing).
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 6134 BP; 1913 A; 76 C; 1273 G; 2872 T; 0 other;

  Query Match             15.2%;  Score 36.6;  DB 22;  Length 6134;
  Best Local Similarity   48.3%;  Pred. No. 0.5;
  Matches  102;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy        2 ATCTGAATCCGACCAACCGTTCTATGAAAATCGTTGTATCTAGACTTAGGCTGGTTGGCA 61

Db     3969 ATTTAAATCCTAACACCTTTATTAACGTTAAATTTATAAATAATCTTAAAACGCGTACCA 3910

Qy       62 AGATACTTTTAGCAACATACTCCGAAACCACTGGACGGATTTTTATGATGTTTGTTTTAG 121

Db     3909 AACTTCTCTAAACTTTAAACCCTTCAACCAATAAAATAATAATAACTTACATAATTACCG 3850

Qy      122 AGGCTTTGGTGACCTGCCTAAAAATACTACAAACAAAATGATTATTTGCGAGAGCGTGGG 181

Db     3849 TAAAAATAAATAACATCGTATATATATTAAATACTTAATAAATTTTAACTTAATAATTAA 3790

Qy      182 CGTTAATATAAAACAAGAACTAATAAACGCT 212

Db     3789 TATTAATTCAAAACAAATAAAAATAAAAACT 3759


RESULT 4
ABN60970
ID   ABN60970 standard; cDNA; 581 BP.
XX
AC   ABN60970;
XX
DT   28-JUN-2002  (first entry)
XX
DE   Human cancer related polynucleotide SEQ ID NO 937.
XX
KW   Human; cytostatic; gene expression; gene mapping; tissue profiling;
KW   gene therapy; cancer; tumour; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200214500-A2.
XX
PD   21-FEB-2002.
XX
PF   16-AUG-2001; 2001WO-US25840.
XX
PR   16-AUG-2000; 2000US-226326P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Escobedo J,  Garcia PD,  Sudduth-Klinger J,  Reinhard C,  Randazzo F;
PI   Lamson G,  Scott EM,  Zhang G,  Kassam A,  Pot D,  Labat I;
XX
DR   WPI; 2002-241905/29.
XX
PT   New nucleic acid for producing a polypeptide, detecting differentially
PT   expressed genes correlated with a cancerous state of a mammalian cell,
PT   and inhibiting tumor growth -
XX
PS   Claim 1; SEQ ID NO 937; 883pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated polynucleotide (ABN27253-ABN33262)
CC   with cytostatic activity. The polynucleotide is used to produce a
CC   polypeptide, to detect differentially expressed genes correlated with a
CC   cancerous state of a mammalian cell and to inhibit tumour growth. The
CC   polynucleotide is used as a probe in mapping and tissue profiling. The
CC   encoded polypeptide and antibodies to the polypeptide can also be used
CC   for therapeutic and diagnostic purposes. The polynucleotide is useful for
CC   gene therapy.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 581 BP; 263 A; 84 C; 95 G; 139 T; 0 other;

  Query Match             14.3%;  Score 34.2;  DB 24;  Length 581;
  Best Local Similarity   47.8%;  Pred. No. 1.4;
  Matches   99;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy       27 GAAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATACTCCGA 86

Db      375 GAAAATCAGTAAAGACATAGTTGAACTGAATAGCACCATTAATAAAGTAGGTCTCATTGA 434

Qy       87 AACCACTGGACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAAT 146

Db      435 CATTTATAGAATATACCAAACCTAGGTTATAAAACAGACATTAATAAATTTAAAAAAAGG 494

Qy      147 ACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAACTAATA 206

Db      495 AAAGCATACGAAATATGCTCTTAGACTACAGTGGAATTTATTGAAAATTAATAACAAATT 554

Qy      207 AACGCTCTCGCACCCGCAATTATATTT 233

Db      555 ATAGCTGTAGAAACCCCCAAAATATTT 581


RESULT 5
ABL17579/c
ID   ABL17579 standard; DNA; 1396 BP.
XX
AC   ABL17579;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 4210.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
```

```
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 4210; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABS57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1396 BP; 365 A; 346 C; 388 G; 297 T; 0 other;

  Query Match          14.2%;  Score 34;  DB 23;  Length 1396;
  Best Local Similarity   57.5%;  Pred. No. 2;
  Matches   61;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy       51 GCTGGTTGGCAAGATACTTTAGCAACATACTCCGAAACCACTGGACGGATTTTTATGAT 110
            ||| ||| |  |||| |   |  | || |  | || | | ||||  ||| | | ||
Db      212 GCTCGTTTTCCGATATTCGCCGATAGCTAATTAGCAAACAATCGACGGCATTTATTTAT 153

Qy      111 GTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAACA 156
            | ||||  |   |  ||||  | |  ||  | || ||| ||| |||
Db      152 GCTTGTAATTTCGCATTTGCCGCCGCGCACACAAGTACCACACACA 107


RESULT 6
ABL17578/c
ID   ABL17578 standard; DNA; 7729 BP.
XX
AC   ABL17578;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 4207.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO2001071042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 4207; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABS57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 7729 BP; 2361 A; 1594 C; 1587 G; 2187 T; 0 other;

  Query Match          14.2%;  Score 34;  DB 23;  Length 7729;
  Best Local Similarity   57.5%;  Pred. No. 3.2;
  Matches   61;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy       51 GCTGGTTGGCAAGATACTTTAGCAACATACTCCGAAACCACTGGACGGATTTTTATGAT 110
            ||| ||| |  |||| |   |  | || |  | || | | ||||  ||| | | ||
Db     1212 GCTCGTTTTCCGATATTCGCCGATAGCTAATTAGCAAACAATCGACGGCATTTATTTAT 1153

Qy      111 GTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAACA 156
            | ||||  |   |  ||||  | |  ||  | || ||| ||| |||
Db     1152 GCTTGTAATTTCGCATTTGCCGCCGCGCACACAAGTACCACACACA 1107


RESULT 7
AAQ75857
ID   AAQ75857 standard; DNA; 1020 BP.
XX
AC   AAQ75857;
XX
DT   06-JUL-1995  (first entry)
XX
DE   Haemolysin regulator DNA.
XX
KW   Haemolysin regulator; tuberculosis; vaccine;
KW   antisense; therapy; ds.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             99..1020
FT                   /*tag= a
FT                   /note= "haemolysin regulator"
FT   misc_feature    99..101
FT                   /*tag= b
FT                   /note= "translation start codon"
FT   -10_signal      76..81
FT                   /*tag= c
FT                   /note= "5/6 match with E. coli promoter -10
FT                    consensus sequence (TATAAT)"
FT   -35_signal      47..52
FT                   /*tag= d
FT                   /note= "4/6 match with -35 consensus sequence
FT                    (TTGACA)"
FT   RBS             87..94
FT                   /*tag= e
FT                   /note= "4/8 match with RBA binding site
FT                    (AGGAAAGG)"
XX
PN   WO9428137-A.
XX
PD   08-DEC-1994.
XX
PF   25-MAY-1994;    94WO-GB05869.
XX
PR   24-MAY-1993;    93US-0066830.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
```

```
XX
PI   King CH,  Sathish M,  Shinnick TM;
XX
DR   WPI; 1995-022809/03.
DR   P-PSDB; AAR67645.
XX
PT   DNA encoding carbapenem R gene product - used for production of
PT   carbapenem antibiotics in bacteria and for bio:assay.
XX
PS   Claim 1; Page 38-39; 46pp; English.
XX
CC   The sequence encodes the E. coli haemolysis protein regulator gene
CC   which may be used, in an avirulent Mycobacterium bovis BCG vaccine
CC   strain, to generate humoral and cellular responses against
CC   Mycobacterium tuberculosis.  Non-coding sequences may be coupled to
CC   reporter genes to screen for compounds that increase gene
CC   expression.  DNA fragments can be used as probes for antisense
CC   therapy to inhibit haemolysis, and the protein regulator gene
CC   itself may be used as an immunogen for vaccines.
XX
SQ   Sequence 1020 BP; 348 A; 195 C; 217 G; 260 T; 0 other;

  Query Match             14.1%;  Score 33.8;  DB 16;  Length 1020;
  Best Local Similarity   54.4%;  Pred. No. 2.1;
  Matches   68;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        100 ATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAACAAAA 159
              | |||  | ||  ||||   | |||  ||       |      || | ||   |  |    ||
Db         34 AATTTATTAATAGTTGTAAAACAGGAGTTTCATTACAATTTATATATTTAAAGAGGCGAA 93

Qy        160 TGATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAACTAATAAACGCTCTCGCAC 219
              |||||||     || | ||| |   | ||| |     | ||| | | | |||||| ||| |
Db         94 TGATTATGACTGAAATCGTTGCAGATAAAACGGTAGAAGTAGTTAAAAACGCAATCGAAA 153

Qy        220 CCGCA 224
              |||||
Db        154 CCGCA 158


RESULT 8
AAX61476
ID   AAX61476 standard; DNA; 3363 BP.
XX
AC   AAX61476;
XX
DT   19-JUL-1999  (first entry)
XX
DE   B. burgdorferi antigenic protein coding sequence, t12.nt.
XX
KW   Antigenic protein; vaccine; Lyme disease; infection; detection; ss.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9859071-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998;   98WO-US12718.
XX
PR   03-SEP-1997;   97US-0057483.
PR   20-JUN-1997;   97US-0050359.
PR   22-JUL-1997;   97US-0053344.
PR   22-JUL-1997;   97US-0053377.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Choi GH,  Erwin AL,  Hanson MS,  Lathigra R;
XX
DR   WPI; 1999-189980/16.
DR   P-PSDB; AAY19780.
XX
PT   New isolated Borrelia burgdorferi nucleic acids  - used to develop
PT   products for the diagnosis, prevention and treatment of diseases
PT   caused by Borrelia, particularly Lyme disease
XX
PS   Claim 1; Page 58-59; 275pp; English.
XX
CC   This sequence encodes a Borrelia burgdorferi (Bb) protein of the
CC   invention, which is suitable for use in a vaccine. The Bb polypeptides
CC   can be used in vaccines for eliciting protective antibodies to members of
CC   the Borrelia genus, particularly for the use against Lyme disease in
CC   humans and animals. They can be used for preventing or attenuating an
CC   infection caused by a member of the Borrelia genus. The products can also
CC   be used for detection of members of the Borrelia genus.
XX
SQ   Sequence 3363 BP; 1193 A; 388 C; 505 G; 1277 T; 0 other;

  Query Match             14.0%;  Score 33.6;  DB 20;  Length 3363;
  Best Local Similarity   50.0%;  Pred. No. 3.4;
  Matches   84;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy         51 GCTGGTTGGCAAGATACTTTTAGCAACATACTCCGAAACCACTGGACGGATTTTTATGAT 110
              || | |     |   || | ||| |    |     |||      ||  |      |   ||| |||
Db        259 GCAGATAATCTTGAAAATTTCACTTATGGACTTACTAAAGATGAAAGTATTATTTTTGAA 318

Qy        111 GTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAACAAAATGATTATTTGC 170
              |    |  |||     ||| | |||       |||| | | || |||  | |   ||||    ||
Db        319 GGAAGAGTTAATATCTTGGTTGAAGATATTAAAGAAAATAAAAAGCACAATATTAAAGGC 378

Qy        171 GAGAGCGTGGGCGTTAATATAAAACAAGAACTAATAAACGCTCTCGCA 218
              || ||  | | | ||||||  ||       ||  |  |   | ||| | | |
Db        379 GACAGAATAGTCCTTAATAAGAACTCTAAAAAACTTTATGCTATTGGA 426


RESULT 9
AAX61475
ID   AAX61475 standard; DNA; 3441 BP.
XX
AC   AAX61475;
XX
DT   19-JUL-1999  (first entry)
XX
DE   B. burgdorferi antigenic protein coding sequence, f12.nt.
XX
KW   Antigenic protein; vaccine; Lyme disease; infection; detection; ss.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9859071-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998;   98WO-US12718.
XX
PR   03-SEP-1997;   97US-0057483.
PR   20-JUN-1997;   97US-0050359.
PR   22-JUL-1997;   97US-0053344.
PR   22-JUL-1997;   97US-0053377.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Choi GH,  Erwin AL,  Hanson MS,  Lathigra R;
XX
DR   WPI; 1999-189980/16.
DR   P-PSDB; AAY19779.
XX
PT   New isolated Borrelia burgdorferi nucleic acids  - used to develop
PT   products for the diagnosis, prevention and treatment of diseases
PT   caused by Borrelia, particularly Lyme disease
XX
PS   Claim 1; Page 57-58; 275pp; English.
XX
CC   This sequence encodes a Borrelia burgdorferi (Bb) protein of the
CC   invention, which is suitable for use in a vaccine. The Bb polypeptides
```

```
CC   can be used in vaccines for eliciting protective antibodies to members of
CC   the Borrelia genus, particularly for the use against Lyme disease in
CC   humans and animals. They can be used for preventing or attenuating an
CC   infection caused by a member of the Borrelia genus. The products can also
CC   be used for detection of members of the Borrelia genus.
XX
SQ   Sequence 3441 BP; 1219 A; 396 C; 513 G; 1313 T; 0 other;

  Query Match             14.0%;  Score 33.6;  DB 20;  Length 3441;
  Best Local Similarity   50.0%;  Pred. No. 3.4;
  Matches   84;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy         51 GCTGGTTGGCAAGATACTTTTAGCAACATACTCCGAAACCACTGGACGGATTTTTATGAT 110
              || | |   |  || | ||| |   |   |||    ||  |    |     | ||| |||
Db        337 GCAGATAATCTTGAAAATTTCACTTATGGACTTACTAAAGATGAAAGTATTATTTTTGAA 396

Qy        111 GTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAACAAAATGATTATTTGC 170
               |   |  |||    ||| | |||       |||| |  | || |||  | |  ||||   ||
Db        397 GGAAGAGTTAATATCTTGGTTGAAGATATTAAAGAAAATAAAAAGCACAATATTAAAGGC 456

Qy        171 GAGAGCGTGGGCGTTAATATAAAACAAGAACTAATAAACGCTCTCGCA 218
              || ||  | | | |||||| ||       ||  | |   | ||| | | |
Db        457 GACAGAATAGTCCTTAATAAGAACTCTAAAAAACTTTATGCTATTGGA 504


RESULT 10
AAD14603
ID   AAD14603 standard; cDNA; 10723 BP.
XX
AC   AAD14603;
XX
DT   01-NOV-2001  (first entry)
XX
DE   Dengue virus (DEN)-2/1-VP chimeric cDNA.
XX
KW   Flavivirus; Dengue virus; DEN; vaccine; infection; virucidal;
KW   avirulent; immunogenic; viral disease; pharmaceutical; chimeric; ss.
XX
OS   Chimeric - Dengue virus type I.
OS   Chimeric - Dengue virus type II.
XX
FH   Key             Location/Qualifiers
FT   CDS             97..10272
FT                   /*tag=  a
FT                   /product= "DEN-2/1-VP fusion protein"
XX
PN   WO200160847-A2.
XX
PD   23-AUG-2001.
XX
PF   16-FEB-2001; 2001WO-US05142.
XX
PR   16-FEB-2000; 2000US-0182829.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Kinney RM,  Kinney CYH,  Butrapet S,  Gubler DL,  Bhamarapravati N;
XX
DR   WPI; 2001-497162/54.
DR   P-PSDB; AAE07982.
XX
PT   Chimeric flaviviruses that are avirulent and immunogenic, useful for
PT   vaccinating against a range of dengue viruses -
XX
PS   Example 1; Page 155-170; 470pp; English.
XX
CC   The invention relates to avirulent, immunogenic flavivirus chimeras
CC   comprising amino acid mutations in the non-structural proteins of a
CC   flavivirus. Chimeric viruses containing the attenuation-mutated non-
CC   structural genes of the virus are used as a backbone into which the
CC   structural protein genes of a second flavivirus strain are inserted.
CC   These chimeric viruses elicit pronounced immunogenicity but lack the
CC   accompanying clinical symptoms of viral disease. Attenuated chimeric
CC   flaviviruses are combined in a pharmaceutical composition to confer
CC   simultaneous immunity against several strains of pathogenic flaviviruses
CC   such as dengue virus serotypes DEN-1, DEN-2, DEN-3 and DEN-4. Immunogenic
CC   flavivirus chimeras are also used as immunogens or multivalent vaccines
CC   to confer simultaneous protection against infections. The present cDNA
CC   sequence encodes dengue virus (DEN)-2/1-VP fusion protein related to the
CC   invention. This fusion protein contains attenuated DEN-2 PDK-53V backbone
CC   comprising a valine at the non-structural protein (NS3)-250 and the
CC   capsid protein (C), premembrane/membrane protein (prM) and an envelope
CC   protein (E) from wild-type DEN-1 16007 virus.
XX
SQ   Sequence 10723 BP; 3514 A; 2229 C; 2722 G; 2258 T; 0 other;

  Query Match             14.0%;  Score 33.6;  DB 22;  Length 10723;
  Best Local Similarity   50.0%;  Pred. No. 4.6;
  Matches   84;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy         27 GAAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATACTCCGA 86
              ||| || |    || || |    ||    | | | |   |||| | | | |   |  ||    |
Db        153 GAGAAACCGCGTGTCAACTGTTTCACAGTTGGCGAAGAGATTCTCAAAAGGATTGCTCTC 212

Qy         87 AACCACTGGACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAAT 146
              |  |   ||||  ||      || || |  | |||      || ||  | |  | . |||   ||
Db        213 AGGCCAAGGACCCATGAAATTGGTGATGGCTTTCATAGCATTCTTAAGATTTCTAGCCAT 272

Qy        147 ACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAA 194
              ||   | ||||  | ||  | ||  || |||   |  |  | ||  |     | ||
Db        273 ACCCCCAACAGCAGGAATTTTGGCCAGATGGGGCTCATTCAAGAAGAA 320


RESULT 11
AAD14604
ID   AAD14604 standard; cDNA; 10723 BP.
XX
AC   AAD14604;
XX
DT   01-NOV-2001  (first entry)
XX
DE   Dengue virus (DEN)-2/1-VV chimeric cDNA.
XX
KW   Flavivirus; Dengue virus; DEN; vaccine; infection; virucidal;
KW   avirulent; immunogenic; viral disease; pharmaceutical; chimeric; ss.
XX
OS   Chimeric - Dengue virus type I.
OS   Chimeric - Dengue virus type II.
XX
FH   Key             Location/Qualifiers
FT   CDS             97..10272
FT                   /*tag=  a
FT                   /product= "DEN-2/1-VV fusion protein"
XX
PN   WO200160847-A2.
XX
PD   23-AUG-2001.
XX
PF   16-FEB-2001; 2001WO-US05142.
XX
PR   16-FEB-2000; 2000US-0182829.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Kinney RM,  Kinney CYH,  Butrapet S,  Gubler DL,  Bhamarapravati N;
XX
DR   WPI; 2001-497162/54.
DR   P-PSDB; AAE07983.
XX
PT   Chimeric flaviviruses that are avirulent and immunogenic, useful for
PT   vaccinating against a range of dengue viruses -
XX
PS   Example 1; Page 179-195; 470pp; English.
XX
CC   The invention relates to avirulent, immunogenic flavivirus chimeras
CC   comprising amino acid mutations in the non-structural proteins of a
```

```
CC   flavivirus. Chimeric viruses containing the attenuation-mutated non-
CC   structural genes of the virus are used as a backbone into which the
CC   structural protein genes of a second flavivirus strain are inserted.
CC   These chimeric viruses elicit pronounced immunogenicity but lack the
CC   accompanying clinical symptoms of viral disease. Attenuated chimeric
CC   flaviviruses are combined in a pharmaceutical composition to confer
CC   simultaneous immunity against several strains of pathogenic flaviviruses
CC   such as dengue virus serotypes DEN-1, DEN-2, DEN-3 and DEN-4. Immunogenic
CC   flavivirus chimeras are also used as immunogens or multivalent vaccines
CC   to confer simultaneous protection against infections. The present cDNA
CC   sequence encodes dengue virus (DEN)-2/1-VV fusion protein related to the
CC   invention. This fusion protein contains attenuated DEN-2 PDK-53V backbone
CC   comprising a valine at the non-structural protein (NS3)-250 and the
CC   capsid protein (C), premembrane/membrane protein (prM) and an envelope
CC   protein (E) from the vaccine strain of Dengue-1 (DEN-1) PDK-13 virus.
XX
SQ   Sequence 10723 BP; 3514 A; 2228 C; 2722 G; 2259 T; 0 other;

  Query Match             14.0%;  Score 33.6;  DB 22;  Length 10723;
  Best Local Similarity   50.0%;  Pred. No. 4.6;
  Matches   84;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy      27 GAAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTCAAAGAGATCTTCCTCGA 86
           || | |   || | |   ||   | | | |||| | | | |   | ||    |
Db     153 GAGAAACCGCGTGTCAACTGTTCACAGTTGGCGAAGAGATCTCAAAAGGATTGCTCTC 212

Qy      87 AACCACTGGACGGATTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAAT 146
           | |   ||||  ||    || | | | |||    || ||   | |  | |||   ||
Db     213 AGGCCAAGGACCCATGAAATTGGTGATGGCTTTCATAGCATTCTTAAGATTTCTAGCCAT 272

Qy     147 ACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAA 194
           ||  | ||||   | ||  | ||  || |||  |  |  | ||  |    ||
Db     273 ACCCCCAACAGCAGGAATTTTGGCTAGATGGGGCTCATTCAAGAAGAA 320


RESULT 12
AAD14601
ID   AAD14601 standard; cDNA; 10735 BP.
XX
AC   AAD14601;
XX
DT   01-NOV-2001  (first entry)
XX
DE   Wild-type, virulent DEN-1 16007 cDNA.
XX
KW   Flavivirus; Dengue virus-1; DEN-1; vaccine; infection; virucidal;
KW   avirulent; immunogenic; viral disease; pharmaceutical; ss.
XX
OS   Dengue virus type I.
XX
FH   Key             Location/Qualifiers
FT   CDS             95..10273
FT                   /*tag= a
FT                   /product= "DEN-1 16007 protein"
XX
PN   WO200160847-A2.
XX
PD   23-AUG-2001.
XX
PF   16-FEB-2001; 2001WO-US05142.
XX
PR   16-FEB-2000; 2000US-0182829.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Kinney RM, Kinney CYH, Butrapet S, Gubler DL, Bhamarapravati N;
XX
DR   WPI; 2001-497162/54.
DR   P-PSDB; AAE07980.
XX
PT   Chimeric flaviviruses that are avirulent and immunogenic, useful for
PT   vaccinating against a range of dengue viruses -
XX

PS   Example 1; Page 106-122; 470pp; English.
XX
CC   The invention relates to avirulent, immunogenic flavivirus chimeras
CC   comprising amino acid mutations in the non-structural proteins of a
CC   flavivirus. Chimeric viruses containing the attenuation-mutated non-
CC   structural genes of the virus are used as a backbone into which the
CC   structural protein genes of a second flavivirus strain are inserted.
CC   These chimeric viruses elicit pronounced immunogenicity but lack the
CC   accompanying clinical symptoms of viral disease. Attenuated chimeric
CC   flaviviruses are combined in a pharmaceutical composition to confer
CC   simultaneous immunity against several strains of pathogenic flaviviruses
CC   such as dengue virus serotypes DEN-1, DEN-2, DEN-3 and DEN-4. Immunogenic
CC   flavivirus chimeras are also used as immunogens or multivalent vaccines
CC   to confer simultaneous protection against infections. The present cDNA
CC   sequence encodes wild-type, virulent dengue-1 (DEN-1) 16007 virus protein
CC   used for constructing flavivirus chimeras. Dengue virus types 1-4 (DEN-1
CC   to DEN-4) are mosquito-borne flavivirus pathogens. The flavivirus genome
CC   contains 5' non-coding region followed by a capsid protein (C) encoding
CC   region, presemb rane/membrane protein (prM) encoding region, an envelope
CC   protein (E) encoding region, followed by the region encoding non-
CC   structural proteins (NS1-NS2A-NS2A-NS3-NS4A-NS4B-NS5) and finally a 3'
CC   non-coding region.
XX
SQ   Sequence 10735 BP; 3432 A; 2232 C; 2774 G; 2297 T; 0 other;

  Query Match             14.0%;  Score 33.6;  DB 22;  Length 10735;
  Best Local Similarity   50.0%;  Pred. No. 4.6;
  Matches   84;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy      27 GAAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATACTCCGA 86
           || | |   || | |   ||   | | | |||| | | | |   | ||    |
Db     151 GAGAAACCGCGTGTCAACTGTTTCACAGTTGGCGAAGAGATTCTCAAAAGGATTGCTCTC 210

Qy      87 AACCACTGGACGGATTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAAT 146
           | |   ||||  ||    || | | | |||    || ||   | |  | |||   ||
Db     211 AGGCCAAGGACCCATGAAATTGGTGATGGCTTTCATAGCATTCTTAAGATTTCTAGCCAT 270

Qy     147 ACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAA 194
           ||  | ||||   | ||  | ||  || |||  |  |  | ||  |    ||
Db     271 ACCCCCAACAGCAGGAATTTTGGCTAGATGGGGCTCATTCAAGAAGAA 318


RESULT 13
AAD14602
ID   AAD14602 standard; cDNA; 10735 BP.
XX
AC   AAD14602;
XX
DT   01-NOV-2001  (first entry)
XX
DE   Attenuated, vaccine-strain DEN-1 PDK-13 variant cDNA.
XX
KW   Flavivirus; Dengue virus-1; DEN-1; vaccine; infection; virucidal; mutein;
KW   avirulent; immunogenic; viral disease; pharmaceutical; mutant; variant;
KW   ss.
XX
OS   Dengue virus type I.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             95..10273
FT                   /*tag= a
FT                   /product= "DEN-1 PDK-13 protein variant"
FT   mutation        replace (1323, T)
FT                   /*tag= b
FT   mutation        replace (1541, G)
FT                   /*tag= c
FT   mutation        replace (1543, A)
FT                   /*tag= d
FT   mutation        replace (1545, G)
FT                   /*tag= e
FT   mutation        replace (1567, A)
FT                   /*tag= f
```

```
FT   mutation        replace (1608, C)
FT                   /*tag= g
FT   mutation        replace (2363, A)
FT                   /*tag= h
FT   mutation        replace (2695, T)
FT                   /*tag= i
FT   mutation        replace (2782, C)
FT                   /*tag= j
FT   mutation        replace (5063, G)
FT                   /*tag= k
FT   mutation        replace (6048, A)
FT                   /*tag= l
FT   mutation        replace (6806, A)
FT                   /*tag= m
FT   mutation        replace (7330, A)
FT                   /*tag= n
FT   mutation        replace (9445, C)
FT                   /*tag= o
XX
PN   WO200160847-A2.
XX
PD   23-AUG-2001.
XX
PF   16-FEB-2001; 2001WO-US05142.
XX
PR   16-FEB-2000; 2000US-0182829.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Kinney RM, Kinney CYH, Butrapet S, Gubler DL, Bhamarapravati N;
XX
DR   WPI; 2001-497162/54.
DR   P-PSDB; AAE07981.
XX
PT   Chimeric flaviviruses that are avirulent and immunogenic, useful for
PT   vaccinating against a range of dengue viruses -
XX
PS   Example 1; Page 130-146; 470pp; English.
XX
CC   The invention relates to avirulent, immunogenic flavivirus chimeras
CC   comprising amino acid mutations in the non-structural proteins of a
CC   flavivirus. Chimeric viruses containing the attenuation-mutated non-
CC   structural genes of the virus are used as a backbone into which the
CC   structural.protein genes of a second flavivirus strain are inserted.
CC   These chimeric viruses elicit pronounced immunogenicity but lack the
CC   accompanying clinical symptoms of viral disease. Attenuated chimeric
CC   flaviviruses are combined in a pharmaceutical composition to confer
CC   simultaneous immunity against several strains of pathogenic flaviviruses
CC   such as dengue virus serotypes DEN-1, DEN-2, DEN-3 and DEN-4. Immunogenic
CC   flavivirus chimeras are also used as immunogens or multivalent vaccines
CC   to confer simultaneous protection against infections. The present cDNA
CC   sequence encodes attenuated dengue-1 (DEN-1) PDK-13 virus protein variant
CC   used for constructing flavivirus chimeras. Dengue virus types 1-4 (DEN-1
CC   to DEN-4) are mosquito-borne flavivirus pathogens. The flavivirus genome
CC   contains 5' non-coding region followed by a capsid protein (C) encoding
CC   region, premembrane/membrane protein (prM) encoding region, an envelope
CC   protein (E) encoding region, followed by the region encoding non-
CC   structural proteins (NS1-NS2A-NS2A-NS3-NS4A-NS4B-NS5) and finally a 3'
CC   non-coding region.
XX
SQ   Sequence 10735 BP; 3429 A; 2231 C; 2776 G; 2299 T; 0 other;

  Query Match             14.0%;  Score 33.6;  DB 22;  Length 10735;
  Best Local Similarity   50.0%;  Pred. No. 4.6;
  Matches   84;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy         27 GAAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATACTCCGA 86
                || || |    || || |    ||     | | | |   |||| | | | |   |   ||    |
Db        151 GAGAAACCGCGTGTCAACTGTTTCACAGTTGGCGAAGAGATTCTCAAAAGGATTGCTCTC 210

Qy         87 AACCACTGGACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAAT 146
              |   |     ||||   ||      || || | | |||       || ||   | |    |   |||    ||
Db        211 AGGCCAAGGACCCATGAAATTGGTGATGGCTTTCATAGCATTCTTAAGATTTCTAGCCAT 270

Qy        147 ACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAA 194
              ||   |  ||||   | ||  | || || |||    | |   | ||  |      | ||
Db        271 ACCCCCAACAGCAGGAATTTTGGCTAGATGGGGCTCATTCAAGAAGAA 318


RESULT 14
ABN70879/c
ID   ABN70879 standard; DNA; 1860 BP.
XX
AC   ABN70879;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Streptococcus polynucleotide SEQ ID NO 9671.
XX
KW   Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW   group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW   antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS   Streptococcus agalactiae.
XX
PN   WO200234771-A2.
XX
PD   02-MAY-2002.
XX
PF   29-OCT-2001; 2001WO-GB04789.
XX
PR   27-OCT-2000; 2000GB-0026333.
PR   24-NOV-2000; 2000GB-0028727.
PR   07-MAR-2001; 2001GB-0005640.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
PI   Tettelin H;
XX
DR   WPI; 2002-352536/38.
DR   P-PSDB; ABP30248.
XX
PT   New Streptococcus protein for the treatment or prevention of infection
PT   or disease caused by Streptococcus bacteria, such as meningitis, and
PT   for detecting a compound that binds to the protein -
XX
PS   Claim 7; Page 4093; 4525pp; English.
XX
CC   The invention relates to a protein (ABP25413-ABP30895) from group B
CC   streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC   (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC   the specification. The proteins have antibacterial and antiinflammatory
CC   activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC   antibodies that bind (I) are used in the manufacture of medicaments for
CC   the treatment or prevention of infection or disease caused by
CC   Streptococcus bacteria, particularly S. agalactiae and S. pyrogenes.
CC   Nucleic acids encoding (I) are used to detect streptococcus in a
CC   biological sample. (I) is used to determine whether a compound binds to
CC   (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC   used as a vaccine or diagnostic composition. The disease caused by
CC   Streptococcus that is prevented or treated may be meningitis. Nucleic
CC   acid encoding (I) may be used to recombinantly produce (I) and may be
CC   used in gene therapy. Antibodies to (I) are used for affinity
CC   chromatography, immunoassays, and distinguishing/identifying
CC   Streptococcus proteins.
XX
SQ   Sequence 1860 BP; 619 A; 296 C; 407 G; 538 T; 0 other;

  Query Match             13.9%;  Score 33.4;  DB 24;  Length 1860;
  Best Local Similarity   52.5%;  Pred. No. 3.3;
  Matches   73;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy         35 TTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATACTCCGAAACCACTG 94
              |||    | | ||           | | || ||| || | |||  ||||| || |   | | |||
```

```
Db   1099 TTGCTGCAATACGGGCTTTAGATGCCAAAATTCCTTTTTCAACAATCTTCTTAGCTACTT 1040

Qy     95 GACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAA 154
          |  |  ||| ||  | |   |    ||  | |||| | | |   | | |     || |   ||
Db   1039 GTGGATTTTCTAATAGGAATCGATTGAAAGCTTCGCTAAAGAGACGATTTGTAATTTTAA 980

Qy    155 CAAAATGATTATTTGCGAG 173
          |||  | |  |||| | ||
Db    979 CAACTTCAGAATTTCCAAG 961
```

```
RESULT 15
ABN66701/c
ID   ABN66701 standard; DNA; 1950 BP.
XX
AC   ABN66701;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Streptococcus polynucleotide SEQ ID NO 1315.
XX
KW   Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW   group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW   antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS   Streptococcus agalactiae.
XX
PN   WO200234771-A2.
XX
PD   02-MAY-2002.
XX
PF   29-OCT-2001; 2001WO-GB04789.
XX
PR   27-OCT-2000; 2000GB-0026333.
PR   24-NOV-2000; 2000GB-0028727.
PR   07-MAR-2001; 2001GB-0005640.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Telford J,  Masignani V,  Margarit Ros YI,  Grandi G,  Fraser C;
PI   Tettelin H;
XX
DR   WPI; 2002-352536/38.
DR   P-PSDB; ABP26070.
XX
PT   New Streptococcus protein for the treatment or prevention of infection
PT   or disease caused by Streptococcus bacteria, such as meningitis, and
PT   for detecting a compound that binds to the protein -
XX
PS   Claim 7; Page 3287-3288; 4525pp; English.
XX
CC   The invention relates to a protein (ABP25413-ABP30895) from group B
CC   streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC   (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC   the specification. The proteins have antibacterial and antiinflammatory
CC   activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC   antibodies that bind (I) are used in the manufacture of medicaments for
CC   the treatment or prevention of infection or disease caused by
CC   Streptococcus bacteria, particularly S. agalactiae and S. pyrogenes.
CC   Nucleic acids encoding (I) are used to detect Streptococcus in a
CC   biological sample. (I) is used to determine whether a compound binds to
CC   (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC   used as a vaccine or diagnostic composition. The disease caused by
CC   Streptococcus that is prevented or treated may be meningitis. Nucleic
CC   acid encoding (I) may be used to recombinantly produce (I) and may be
CC   used in gene therapy. Antibodies to (I) are used for affinity
CC   chromatography, immunoassays, and distinguishing/identifying
CC   Streptococcus proteins.
XX
SQ   Sequence 1950 BP; 656 A; 309 C; 428 G; 557 T; 0 other;

  Query Match             13.9%;  Score 33.4;  DB 24;  Length 1950;
  Best Local Similarity   52.5%;  Pred. No. 3.4;
  Matches   73;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy     35 TTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATACTCCGAAACCACTG 94
          |||   | | ||       | | || ||| || | |||  |||||  || |   | | |||
Db   1189 TTGCTGCAATACGGGCTTTAGATGCCAAAATTCCTTTTTCAACAATCTTCTTAGCTACTT 1130

Qy     95 GACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAA 154
          |  |  ||| ||  | |   |    ||  | |||| | | |   | | |     || |   ||
Db   1129 GTGGATTTTCTAATAGGAATCGATTGAAAGCTTCGCTAAAGAGACGATTTGTAATTTTAA 1070

Qy    155 CAAAATGATTATTTGCGAG 173
          |||  | |  |||| | ||
Db   1069 CAACTTCAGAATTTCCAAG 1051

Search completed: May 12, 2003, 01:14:51
Job time : 434.571 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 01:08:23 ; Search time 105.429 Seconds
(without alignments)
698.126 Million cell updates/sec

Title: US-09-980-277-1
Perfect score: 240
Sequence: 1 gatctgaatccgaccaaccg.....cgcaattatattttgttctt 240

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	33.8	14.1	1020	1	US-08-557-115-2	Sequence 2, Appli
2	33.8	14.1	1020	5	PCT-US94-05869-2	Sequence 2, Appli
3	30.4	12.7	10718	3	US-08-325-426B-1	Sequence 1, Appli
4	28.8	12.0	3047	1	US-08-280-690-1	Sequence 1, Appli
c 5	28.6	11.9	1271	1	US-08-464-339A-1	Sequence 1, Appli
c 6	28.6	11.9	1271	5	PCT-US94-14388-1	Sequence 1, Appli
c 7	28.6	11.9	40138	3	US-09-090-793-12	Sequence 12, Appl
8	28.6	11.9	162450	4	US-09-442-882-1	Sequence 1, Appli
c 9	28.4	11.8	943	4	US-09-641-638-99	Sequence 99, Appl
10	28.2	11.8	3785	4	US-09-889-718-1	Sequence 1, Appli
c 11	28.2	11.8	7351	1	US-08-224-391-83	Sequence 83, Appl
c 12	28.2	11.8	7351	1	US-08-484-304-83	Sequence 83, Appl
c 13	28.2	11.8	7351	2	US-08-184-009-127	Sequence 127, App
c 14	28.2	11.8	7351	2	US-08-566-398-39	Sequence 39, Appl
c 15	28.2	11.8	7351	2	US-08-458-356-127	Sequence 127, App
c 16	28.2	11.8	7351	2	US-08-658-665-39	Sequence 39, Appl
c 17	28.2	11.8	7351	4	US-08-796-101-3	Sequence 3, Appli
c 18	28.2	11.8	7351	4	US-08-460-736-127	Sequence 127, App
c 19	28.2	11.8	7351	4	US-09-085-273-39	Sequence 39, Appl
20	28.2	11.8	19011	1	US-08-310-356-36	Sequence 36, Appl
21	28.2	11.8	19557	5	PCT-US92-06300-1	Sequence 1, Appli
22	28	11.7	443	4	US-09-397-787-176	Sequence 176, App
c 23	28	11.7	448	2	US-08-967-101-66	Sequence 66, Appl
c 24	28	11.7	448	2	US-08-592-541-66	Sequence 66, Appl
c 25	28	11.7	448	3	US-09-124-698-66	Sequence 66, Appl
c 26	28	11.7	448	4	US-09-127-480-66	Sequence 66, Appl
c 27	28	11.7	448	4	US-08-496-841C-66	Sequence 66, Appl
c 28	28	11.7	448	4	US-09-124-523-66	Sequence 66, Appl
29	28	11.7	1049	4	US-08-961-527-317	Sequence 317, App
c 30	28	11.7	4184	2	US-08-785-310A-4	Sequence 4, Appli
31	27.8	11.6	1700	2	US-08-897-340-4	Sequence 4, Appli
32	27.8	11.6	1700	3	US-09-252-329-4	Sequence 4, Appli
33	27.6	11.5	1630	4	US-09-175-928-17	Sequence 17, Appl
c 34	27.6	11.5	3418	4	US-09-193-562D-29	Sequence 29, Appl
35	27.6	11.5	7425	4	US-09-453-702B-212	Sequence 212, App
c 36	27.6	11.5	11309	4	US-08-961-527-108	Sequence 108, App
37	27.4	11.4	921	4	US-09-134-001C-1841	Sequence 1841, Ap
38	27.4	11.4	5099	1	US-08-487-890A-4	Sequence 4, Appli
39	27.4	11.4	5099	2	US-08-478-435-4	Sequence 4, Appli
40	27.4	11.4	5099	2	US-08-337-483-4	Sequence 4, Appli
41	27.4	11.4	5099	2	US-08-478-573-4	Sequence 4, Appli
42	27.4	11.4	5099	3	US-08-474-671-4	Sequence 4, Appli
43	27.4	11.4	5099	3	US-08-483-577A-4	Sequence 4, Appli
44	27.4	11.4	5099	4	US-08-897-438-4	Sequence 4, Appli
45	27.4	11.4	5099	4	US-08-637-654-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-557-115-2
; Sequence 2, Application US/08557115
; Patent No. 5731151
; GENERAL INFORMATION:
; APPLICANT: King, Harold C.
; APPLICANT: Sathish, Mundayoor
; APPLICANT: Shinnick, Thomas M.
; TITLE OF INVENTION: Regulator of Contact-Mediated Hemolysin
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones & Askew
; STREET: 191 Peachtree Street, 37th Floor
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.A.
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,115
; FILING DATE: 26-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Greene, Jamie L.
; REGISTRATION NUMBER: 32,467
; REFERENCE/DOCKET NUMBER: 03063-0171US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 818-3700
; TELEFAX: (404) 818-3799
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1020 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: E. coli
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 99..101
; OTHER INFORMATION: /note= "Translation start codon"
; FEATURE:
; NAME/KEY: misc_feature

```
;       LOCATION:  76..81
;       OTHER INFORMATION:  /note= "5/6 match with -10
;       OTHER INFORMATION:  consensus sequence for E. coli promoters (TATAAT)"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  47..52
;       OTHER INFORMATION:  /note= "4/6 match with -35
;       OTHER INFORMATION:  consensus sequence (TTGACA)"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  87..94
;       OTHER INFORMATION:  /note= "4/8 match with the E. coli
;       OTHER INFORMATION:  consensus sequence for ribosome binding sites (AGGAAAGG)"
US-08-557-115-2

  Query Match             14.1%;  Score 33.8;  DB 1;  Length 1020;
  Best Local Similarity   54.4%;  Pred. no. 0.22;
  Matches   68;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy     100 ATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAACAAAA 159
           | ||| | ||  |||| |.||| ||        |    || | ||  |  |    ||
Db      34 AATTTATTAATAGTTGTAAAACAGGAGTTTCATTACAATTTATATATTTAAAGAGGCGAA 93

Qy     160 TGAGTATTTGCGAGAGCGTGGGGCGTTAATATAAAACAAGAACTAATAAACGCTCTCGCAC 219
           |||||||    || | ||| | | ||| |    | ||| | | | |||||| ||| |
Db      94 TGATTATGACTGAAATCGTTGCAGATAAAACGGTAGAAGTAGTTAAAAACGCAATCGAAA 153

Qy     220 CCGCA 224
           |||||
Db     154 CCGCA 158


RESULT 2
PCT-US94-05869-2
; Sequence 2, Application PC/TUS9405869
;  GENERAL INFORMATION:
;    APPLICANT:  King, C. H.
;    APPLICANT:  Sathish, Mundayoor
;    APPLICANT:  Shinnick, Thomas M.
;    TITLE OF INVENTION:  REGULATOR OF CONTACT-MEDIATED HEMOLYSIN
;    TITLE OF INVENTION:  OF MYCOBACTERIUM TUBERCULOSIS
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Suite 1200, The Candler Building
;      STREET:  127 Peachtree Street, N.E.
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303-1811
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/05869
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perryman, David G.
;      REGISTRATION NUMBER:  33,438
;      REFERENCE/DOCKET NUMBER:  1414.611
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 688-0770
;      TELEFAX:  (404) 688-9880
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1020 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  E. coli
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  99..101
;      OTHER INFORMATION:  /note= "Translation start codon"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  76..81
;      OTHER INFORMATION:  /note= "5/6 match with -10
;      OTHER INFORMATION:  consensus sequence for E. coli promoters (TATAAT)"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  47..52
;      OTHER INFORMATION:  /note= "4/6 match with -35
;      OTHER INFORMATION:  consensus sequence (TTGACA)"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  87..94
;      OTHER INFORMATION:  /note= "4/8 match with the E. coli
;      OTHER INFORMATION:  consensus sequence for ribosome binding sites
;      OTHER INFORMATION:  (AGGAAAGG)"
PCT-US94-05869-2

  Query Match             14.1%;  Score 33.8;  DB 5;  Length 1020;
  Best Local Similarity   54.4%;  Pred. no. 0.22;
  Matches   68;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy     100 ATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAACAAAA 159
           | ||| | ||  |||| |.||| ||        |    || | ||  |  |    ||
Db      34 AATTTATTAATAGTTGTAAAACAGGAGTTTCATTACAATTTATATATTTAAAGAGGCGAA 93

Qy     160 TGAGTATTTGCGAGAGCGTGGGGCGTTAATATAAAACAAGAACTAATAAACGCTCTCGCAC 219
           |||||||    || | ||| | | ||| |    | ||| | | | |||||| ||| |
Db      94 TGATTATGACTGAAATCGTTGCAGATAAAACGGTAGAAGTAGTTAAAAACGCAATCGAAA 153

Qy     220 CCGCA 224
           |||||
Db     154 CCGCA 158


RESULT 3
US-08-325-426B-1
; Sequence 1, Application US/08325426B
; Patent No. 6017535
;  GENERAL INFORMATION:
;    APPLICANT:  FU, Jianlin
;    APPLICANT:  TAN, Boon-Haun
;    APPLICANT:  YAP, Eu-Hian
;    APPLICANT:  CHAN, Yow-Cheong
;    APPLICANT:  TAN, Yin-Hwee
;    TITLE OF INVENTION:  cDNA SEQUENCE OF DENGUE VIRUS SEROTYPE 1
;    TITLE OF INVENTION:  (SINGAPORE STRAIN)
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON and VANDERHYE PC
;      STREET:  8th FLOOR, 1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  USA
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/325,426B
;      FILING DATE:  16-DEC-1994
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10718 base pairs
```

```
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  CDNA sequence corresponding to
;    MOLECULE TYPE:  the genomic RNA of DEN1-S275/90
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Dengue Fever Virus Type 1
;      STRAIN:  S275/90
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  81..10268
US-08-325-426B-1

  Query Match              12.7%;  Score 30.4;  DB 3;  Length 10718;
  Best Local Similarity    48.8%;  Pred. No. 4.9;
  Matches   82;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy       27 GAAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATACTCCGA 86
Db      137 GAGAAACCGCGTGTCAACTGGTTCACAGTTGGCGAAGAGATTCTCAAAAGGATTGCTTTC 196

Qy       87 AACCACTGGACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAAT 146
Db      197 AGGCCAAGGACCCATGAAATTGGTGATGGCTTTCATAGCATTCCTAAGATTTCTAGCCAT 256

Qy      147 ACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAA 194
Db      257 ACCCCCAACAGCAGGAATTTTGGCTAGATGGGGCTCATTCAAGAAGAA 304


RESULT 4
US-08-280-690-1
; Sequence 1, Application US/08280690
; Patent No. 5583011
;  GENERAL INFORMATION:
;    APPLICANT:  Mansour, Tag E
;    TITLE OF INVENTION:  Compositions, Treatments, and
;    TITLE OF INVENTION:  Diagnostics for Schistosomiasis and Related Diseases
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooley Godward
;      STREET:  5 Palo Alto
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94306-2155
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/280,690
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Neeley, Richard L
;      REGISTRATION NUMBER:  30,092
;      REFERENCE/DOCKET NUMBER:  STAN-141/00US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-843-5070
;      TELEFAX:  415-857-0663
;      TELEX:  380816 CooleyPA
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3047 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  147..2489
US-08-280-690-1

  Query Match              12.0%;  Score 28.8;  DB 1;  Length 3047;
  Best Local Similarity    49.3%;  Pred. No. 10;
  Matches   75;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy       51 GCTGGTTGGCAAGATACTTTTAGCAACATACTCCGAAACCACTGGACGGATTTTTATGAT 110
Db      861 GATGATTGGCGTGAAAAATTATGTCATAAACTTCGAATGAATCGTGAACATGGACAACGT 920

Qy      111 GTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAACAAAATGATTATTTGC 170
Db      921 GTCAATATAATTATGGTTGCTGAAGGTGCTATCGATAGAGCATGTAAACCGATCACTTGT 980

Qy      171 GAGAGCGTGGGCGTTAATATAAAACAAGAACT 202
Db      981 GAAATAGTGAAAAATTTAATAGTATCAGAACT 1012


RESULT 5
US-08-464-339A-1/c
; Sequence 1, Application US/08464339A
; Patent No. 5747280
;  GENERAL INFORMATION:
;    APPLICANT:  HASTINGS, ET AL.
;    TITLE OF INVENTION:  Human Vascular IBP-Like Growth
;    TITLE OF INVENTION:  Factor
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/464,339A
;      FILING DATE:  June 5, 1995
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/14388
;      FILING DATE:  9 DEC 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MULLINS, J.G.
;      REGISTRATION NUMBER:  33,073
;      REFERENCE/DOCKET NUMBER:  325800-332
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1271 BASE PAIRS
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  cDNA
US-08-464-339A-1

  Query Match              11.9%;  Score 28.6;  DB 1;  Length 1271;
  Best Local Similarity    48.7%;  Pred. No. 9;
  Matches   73;  Conservative    1;  Mismatches   76;  Indels    0;  Gaps    0;

Qy       13 ACCAACCGTTCTATGAAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTA 72
Db     1240 ATCAGCAGCCTTTTGRAATTGCTCTCAGNTCAAAGCTGTTTGTTACTCAAATTTCCATAA 1181
```

```
Qy      73 GCAACATACTCCGAAACCACTGGACGGATTTTATGATGTTTGTTTAGAGGCTTTGGTG 132
           ||  || ||  |  |    ||  | || ||||   | |||||  ||| ||  ||
Db   1180 GCTTCAAACATCTTACTTCCTTCAGGGGTTTTCTGGTTGTTTATTTTGACTTTTCCCAA 1121

Qy     133 ACCTGCCTAAAAATACTACAAACAAAATGA 162
           | |    ||||| |  | || ||  ||| |
Db   1120 AGCCCAAAAAAAAAAAAAAAAGCAYAATTA 1091


RESULT 6
PCT-US94-14388-1/c
; Sequence 1, Application PC/TUS9414388
;  GENERAL INFORMATION:
;     APPLICANT:  HASTINGS, ET AL.
;     TITLE OF INVENTION:  Human Vascular ISP-Like Growth Factor
;     NUMBER OF SEQUENCES:  2
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;       ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;       STREET:  6 BECKER FARM ROAD
;       CITY:  ROSELAND
;       STATE:  NEW JERSEY
;       COUNTRY:  USA
;       ZIP:  07068
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5 INCH DISKETTE
;       COMPUTER:  IBM PS/2
;       OPERATING SYSTEM:  MS-DOS
;       SOFTWARE:  MORD PERFECT 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US94/14388
;       FILING DATE:  Concurrently
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  FERRARO, GREGORY D.
;       REGISTRATION NUMBER:  36,134
;       REFERENCE/DOCKET NUMBER:  325800-219
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  201-994-1700
;       TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1271 BASE PAIRS
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  SINGLE
;       TOPOLOGY:  LINEAR
;     MOLECULE TYPE:  cDNA
PCT-US94-14388-1

  Query Match             11.9%;  Score 28.6;  DB 5;  Length 1271;
  Best Local Similarity   48.7%;  Pred. No. 9;
  Matches   73;  Conservative    1;  Mismatches   76;  Indels    0;  Gaps    0;

Qy      13 ACCAACCGTTCTATGAAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTA 72
           | || | |    | ||:|| |  |   |  |  |   | || |   |||  | |  | |
Db   1240 ATCAGCAGCCTTTTGRAATTGCTCTCAGNTCAAAGCTGTTTGTTACTCAAATTTCCATAA 1181

Qy      73 GCAACATACTCCGAAACCACTGGACGGATTTTATGATGTTTGTTTAGAGGCTTTGGTG 132
           ||  || ||  |  |    ||  | || ||||   | |||||  ||| ||  ||
Db   1180 GCTTCAAACATCTTACTTCCTTCAGGGGTTTTCTGGTTGTTTATTTTGACTTTTCCCAA 1121

Qy     133 ACCTGCCTAAAAATACTACAAACAAAATGA 162
           | |    ||||| |  | || ||  ||| |
Db   1120 AGCCCAAAAAAAAAAAAAAAAGCAYAATTA 1091


RESULT 7
US-09-090-793-12/c
; Sequence 12, Application US/09090793
;
; Patent No. 6140486
; GENERAL INFORMATION:
;  APPLICANT: Calgene, LLC
;  TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression
;  TITLE OF INVENTION: of polyketide-like synthesis genes in plants
;  FILE REFERENCE: CGNE.131.01US
;  CURRENT APPLICATION NUMBER: US/09/090,793
;  CURRENT FILING DATE: 1998-06-04
;  EARLIER APPLICATION NUMBER: 60/048,650
;  EARLIER FILING DATE: 1997-06-04
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;  LENGTH: 40138
;  TYPE: DNA
;  ORGANISM: Vibrio marinus
US-09-090-793-12

  Query Match             11.9%;  Score 28.6;  DB 3;  Length 40138;
  Best Local Similarity   50.4%;  Pred. No. 26;
  Matches   70;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy      96 ACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAAC 155
           ||| || |||| |  |  |  || ||  |||| |  || |  |  ||||  |  |  |
Db  10727 ACGCCTTATTATCAACGTGATTGGAAAGATTTTGATAACGTTGCACAAAATGGTGGTGAT 10668

Qy     156 AAAATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAACTAATAAACGCTCTC 215
           ||||  || ||  |  |  |||  | |  | ||||   || ||||  | | |
Db  10667 AAAAGCGTTGTTAAGCAAGGTTTGGAAGATGCAGTTAAACGTCAATTAATGTGTGATGTA 10608

Qy     216 GCACCCGCAATTATATTTT 234
           |  ||  |  |||| |
Db  10607 CCTTACGGCGTATTATTAT 10589


RESULT 8
US-09-345-882-1
; Sequence 1, Application US/09345882
; Patent No. 6399373
; GENERAL INFORMATION:
;  APPLICANT: Bouqueleret, Lydie
;  TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RI
;  TITLE OF INVENTION:  AND POLYMORPHIC MARKERS ASSOCIATED WITH SAID NUCLEIC ACID.
;  FILE REFERENCE: GENSET.031A
;  CURRENT APPLICATION NUMBER: US/09/345,882
;  CURRENT FILING DATE:  1999-06-30
;  PRIOR APPLICATION NUMBER: US 60/091,315
;  PRIOR FILING DATE:  1998-06-30
;  PRIOR APPLICATION NUMBER: US 60/111,909
;  PRIOR FILING DATE:  1998-12-10
;  NUMBER OF SEQ ID NOS: 140
;  SOFTWARE: Patent.pm
; SEQ ID NO 1
;  LENGTH: 162450
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 72794
;  OTHER INFORMATION: 5-124-273  :  polymorphic  base  A  or  G
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 88073
;  OTHER INFORMATION: 5-127-261  :  polymorphic  base  A  or  C
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 90842
;  OTHER INFORMATION: 99-1437-325  :  polymorphic  base  A  or  G
;  FEATURE:
;  NAME/KEY: allele
;  LOCATION: 93714
;  OTHER INFORMATION: 5-128-60  :  polymorphic  base  deletion  of  GT
;  FEATURE:
```

; NAME/KEY: allele
; LOCATION: 97122
; OTHER INFORMATION: 99-1442-224 : polymorphic base G or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 97152
; OTHER INFORMATION: 5-129-144 : polymorphic base deletion of T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 99098
; OTHER INFORMATION: 5-130-257 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 99117
; OTHER INFORMATION: 5-130-276 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 103806
; OTHER INFORMATION: 5-131-395 : polymorphic base A or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 106940
; OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 108106
; OTHER INFORMATION: 5-135-155 : polymorphic base insertion of A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 108149
; OTHER INFORMATION: 5-135-198 : polymorphic base insertion of GTTT
; FEATURE:
; NAME/KEY: allele
; LOCATION: 108308
; OTHER INFORMATION: 5-135-357 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 108471
; OTHER INFORMATION: 5-136-174 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 134134
; OTHER INFORMATION: 5-140-120 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 134362
; OTHER INFORMATION: 5-140-348 : polymorphic base insertion of A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 134374
; OTHER INFORMATION: 5-140-361 : polymorphic base insertion of CA
; FEATURE:
; NAME/KEY: allele
; LOCATION: 146328
; OTHER INFORMATION: 5-143-84 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 146345
; OTHER INFORMATION: 5-143-101 : polymorphic base A or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 150329
; OTHER INFORMATION: 5-145-24 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 160031
; OTHER INFORMATION: 5-148-352 : polymorphic base G or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 72771..72817
; OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30
; FEATURE:
; NAME/KEY: allele
; LOCATION: 72771..72817
; OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51
; FEATURE:
; NAME/KEY: allele
; LOCATION: 88050..88096
; OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31
; FEATURE:
; NAME/KEY: allele
; LOCATION: 88050..88096
; OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52
; FEATURE:
; NAME/KEY: allele
; LOCATION: 90819..90865
; OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49
; FEATURE:
; NAME/KEY: allele
; LOCATION: 90819..90865
; OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70
; FEATURE:
; NAME/KEY: allele
; LOCATION: 93690..93736
; OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32
; FEATURE:
; NAME/KEY: allele
; LOCATION: 93690..93736
; OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
; FEATURE:
; NAME/KEY: allele
; LOCATION: 97099..97145
; OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50
; FEATURE:
; NAME/KEY: allele
; LOCATION: 97099..97145
; OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71
; FEATURE:
; NAME/KEY: allele
; LOCATION: 97130..97177
; OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
; FEATURE:
; NAME/KEY: allele
; LOCATION: 97130..97177
; OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
; FEATURE:
; NAME/KEY: allele
; LOCATION: 99075..99121
; OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
; FEATURE:
; NAME/KEY: allele
; LOCATION: 99075..99121
; OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
; FEATURE:
; NAME/KEY: allele
; LOCATION: 99094..99140
; OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
; FEATURE:
; NAME/KEY: allele
; LOCATION: 99094..99140
; OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56
; FEATURE:
; NAME/KEY: allele
; LOCATION: 103783..103828
; OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36
; FEATURE:
; NAME/KEY: allele
; LOCATION: 103783..103828
; OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57
; FEATURE:
; NAME/KEY: allele
; LOCATION: 106918..106966
; OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37
; FEATURE:
; NAME/KEY: allele
; LOCATION: 106918..106966

```
;   OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 108084..108130
;   OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 108084..108130
;   OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 108127..108177
;   OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID39
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 108127..108177
;   OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60
;   FEATURE:

  Query Match            11.9%;  Score 28.6;  DB 4;  Length 162450;
  Best Local Similarity  64.2%;  Pred. No. 39;
  Matches   43;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy      142 AAAATACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAAC 201
            |||| || | || ||||||| | ||  || ||   ||     |||||||||    | ||
Db    58404 AAAAAACAAAAACAAAATGAGCAATTCAGAAAGATGGGTAAATAATATAAATAGATAAT 58463

Qy      202 TAATAAA 208
            ||||| |
Db    58464 TAATAGA 58470


RESULT 9
US-09-641-638-99/c
; Sequence 99, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
;  APPLICANT:  Blumenfeld, Marta
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Chumakov, Ilya
;  APPLICANT:  Cohen, Annick
;  TITLE OF INVENTION:  BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
;  TITLE OF INVENTION:  GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
;  FILE REFERENCE:  GENSET.051CP1
;  CURRENT APPLICATION NUMBER: US/09/641,638
;  CURRENT FILING DATE:  2000-08-16
;  PRIOR APPLICATION NUMBER:  US 09/502,330
;  PRIOR FILING DATE:  2000-02-11
;  PRIOR APPLICATION NUMBER:  US 60/133,200
;  PRIOR FILING DATE:  1999-05-07
;  PRIOR APPLICATION NUMBER:  US 09/275,267
;  PRIOR FILING DATE:  1999-03-23
;  PRIOR APPLICATION NUMBER:  US 60/119,917
;  PRIOR FILING DATE:  1999-02-12
;  NUMBER OF SEQ ID NOS:  1304
;  SOFTWARE:  Patent.pm
; SEQ ID NO 99
;   LENGTH: 943
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 443
;   OTHER INFORMATION: 12-386-155 : polymorphic base G or T
;   NAME/KEY: misc_binding
;   LOCATION: 423..442
;   OTHER INFORMATION: 12-386-155.mis1, potential
;   NAME/KEY: misc_binding
;   LOCATION: 444..462
;   OTHER INFORMATION: 12-386-155.mis2, complement
;   NAME/KEY: primer_bind
;   LOCATION: 290..310
;   OTHER INFORMATION: upStream amplification primer
;   NAME/KEY: primer_bind
;   LOCATION: 669..688
;   OTHER INFORMATION: downstream amplification primer, complement
;   NAME/KEY: misc_binding
;   LOCATION: 431..455
;   OTHER INFORMATION: 12-386-155 potential probe
US-09-641-638-99

  Query Match            11.8%;  Score 28.4;  DB 4;  Length 943;
  Best Local Similarity  56.4%;  Pred. No. 9.4;
  Matches   53;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy      141 AAAAATACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAA 200
              | ||||      ||  || |  |    | || || ||  ||||| | |||||||||||  |
Db      938 ATAAATGAATAAATGAAGAAAAAACTCTGGGATAGGATGGGCTTAAATATAAAACAGAAC 879

Qy      201 CTAATAAACGCTCTCGCACCCGCAATTATATTTT 234
              | ||  ||  ||  | ||  |      | ||| | |
Db      878 CCAAATAATACTAGCACATTCCAGTTCATAGTGT 845


RESULT 10
US-09-889-718-1
; Sequence 1, Application US/09889718
; Patent No. 6412816
; GENERAL INFORMATION:
;  APPLICANT: Aventis CropScience GmbH
;  TITLE OF INVENTION: Promotor for gene expression in caryopses of plants
;  FILE REFERENCE: 514413-3886
;  CURRENT APPLICATION NUMBER: US/09/889,718
;  CURRENT FILING DATE:  2001-07-05
;  PRIOR APPLICATION NUMBER: DE 100 41 861.9
;  PRIOR FILING DATE: 2000-08-26
;  PRIOR APPLICATION NUMBER: DE 100 32 379.0
;  PRIOR FILING DATE: 2000-07-06
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 3785
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-09-889-718-1

  Query Match            11.8%;  Score 28.2;  DB 4;  Length 3785;
  Best Local Similarity  52.1%;  Pred. No. 17;
  Matches   63;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy       91 ACTGGACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTA 150
            ||| |     | |||||| ||  ||  || || |  ||| ||       |  || |      |
Db      209 ACTAGTTTCAGTTTATTATTGTTTATTAAGTGTTTTAGTTGGTTTTCTCATTTATGTG 268

Qy      151 CAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAACTAATAAACG 210
              |  ||  |  ||||  | | | | | | | ||||||||| |    |      | ||| ||
Db      269 TATGCATGAATATTAGGGGTGTGTGCGTGTGTTAATATACACATAAGTATTATACACC 328

Qy      211 C 211
            |
Db      329 C 329


RESULT 11
US-08-224-391-83/c
; Sequence 83, Application US/08224391
; Patent No. 5744140
;  GENERAL INFORMATION:
;     APPLICANT:  Paoletti, Enzo
;     APPLICANT:  Pincus, Steven E.
;     TITLE OF INVENTION:  FLAVIVIRUS RECOMBINANT POXVIRUS VACCINE
;     NUMBER OF SEQUENCES:  93
;     CORRESPONDENCE ADDRESS:
;        ADDRESSEE:  Curtis, Morris & Safford
;        ADDRESSEE:  c/o William S. Frommer
;        STREET:  530 Fifth Avenue
```

```
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  United States of America
;       ZIP:  10036
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/224,391
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/729,800
;       FILING DATE:  17-JUL-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Frommer, William S.
;       REGISTRATION NUMBER:  25,506
;       REFERENCE/DOCKET NUMBER:  454310-2340
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 840-3333
;       TELEFAX:  (212) 840-0712
;   INFORMATION FOR SEQ ID NO:  83:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  7351 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-224-391-83

  Query Match           11.8%;  Score 28.2;  DB 1;  Length 7351;
  Best Local Similarity   48.4%;  Pred. No. 20;
  Matches   78;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy      28 AAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATACTCCGAA 87
           | ||   || || |    | |  | | ||| || |  |||| | |   ||| ||
Db    6654 ATAAAGTTTAAATTTTTATCATACCAGGTGGAAACAATCTTTCATCGTTATATTCTTTT 6595

Qy      88 ACCACTGGACGGATTTTATGATGTTTGTTTAGAGGCTTTGGTGACCTGCCTAAAAATA 147
           | | ||   |||| || |||  || ||  |    | |    |    |    | | |
Db    6594 TTAAGTAGATATATTTGTAAGATAATTATTGAATTTACATATTTATTATCAAAATTAGTC 6535

Qy     148 CTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAAT 188
           || ||  || | || |||| ||  || ||  |   || |
Db    6534 ATATAAGTAAGAGGAGTATTTCCGTAAGAGTCAGTAATATT 6494


RESULT 12
US-08-484-304-83/c
; Sequence 83, Application US/08484304
; Patent No. 5744141
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  Pincus, Steven E.
;    TITLE OF INVENTION:  FLAVIVIRUS RECOMBINANT POXVIRUS VACCINE
;    NUMBER OF SEQUENCES:  93
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      ADDRESSEE:  c/o William S. Frommer
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,304
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/224,391
;      FILING DATE:
;      APPLICATION NUMBER:  US 07/729,800
;      FILING DATE:  17-JUL-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2340
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;  INFORMATION FOR SEQ ID NO:  83:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7351 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-484-304-83

  Query Match           11.8%;  Score 28.2;  DB 1;  Length 7351;
  Best Local Similarity   48.4%;  Pred. No. 20;
  Matches   78;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy      28 AAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATACTCCGAA 87
           | ||   || || |    | |  | | ||| || |  |||| | |   ||| ||
Db    6654 ATAAAGTTTAAATTTTTATCATACCAGGTGGAAACAATCTTTCATCGTTATATTCTTTT 6595

Qy      88 ACCACTGGACGGATTTTATGATGTTTGTTTAGAGGCTTTGGTGACCTGCCTAAAAATA 147
           | | ||   |||| || |||  || ||  |    | |    |    |    | | |
Db    6594 TTAAGTAGATATATTTGTAAGATAATTATTGAATTTACATATTTATTATCAAAATTAGTC 6535

Qy     148 CTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAAT 188
           || ||  || | || |||| ||  || ||  |   || |
Db    6534 ATATAAGTAAGAGGAGTATTTCCGTAAGAGTCAGTAATATT 6494


RESULT 13
US-08-184-009-127/c
; Sequence 127, Application US/08184009
; Patent No. 5833975
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  Tartaglia, James
;    APPLICANT:  Cox, William I.
;    TITLE OF INVENTION:  RECOMBINANT VIRUS IMMUNOTHERAPY
;    NUMBER OF SEQUENCES:  217
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/184,009
;      FILING DATE:  19-JAN-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2530
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;      TELEX:  425066CURTMS
;  INFORMATION FOR SEQ ID NO:  127:
```

```
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  7351 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-184-009-127

  Query Match             11.8%;  Score 28.2;  DB 2;  Length 7351;
  Best Local Similarity   48.4%;  Pred. No. 20;
  Matches   78;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy      28 AAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATACTCCGAA 87
             | ||   ||  || |     | |  |  | ||| || |  |||| | |   ||| ||
Db    6654 ATAAAGTTTAAATTTTTATCATACCAGGTGGAAACAATCTTTCATCGTTATATTCTTTT 6595

Qy      88 ACCACTGGACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATA 147
                | | ||   |||| || |||  || ||  |      | |    |     |    | | |
Db    6594 TTAAGTAGATATATTTGTAAGATAATTATTGAATTTACATATTTATTATCAAAATTAGTC 6535

Qy     148 CTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAAT 188
             || ||  || | || ||||| ||  || ||   |    || |
Db    6534 ATATAAGTAAGAGGAGTATTTCCGTAAGAGTCAGTAATATT 6494


RESULT 14
US-08-566-398-39/c
; Sequence 39, Application US/08566398
; Patent No. 5858373
;  GENERAL INFORMATION:
;     APPLICANT:  Paoletti, Enzo
;     APPLICANT:  Gettig, Russell
;     TITLE OF INVENTION:  RECOMBINANT POXVIRUS - FELINE INFECTIOUS
;     TITLE OF INVENTION:  PERITONITIS VIRUS, COMPOSITIONS THEREOF, AND METHODS FOR
;     TITLE OF INVENTION:  MAKING AND USING THEM
;     NUMBER OF SEQUENCES:  63
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Curtis, Morris & Safford, P.C.
;       STREET:  530 Fifth Avenue
;       CITY:  New York
;       STATE:  NY
;       COUNTRY:  USA
;       ZIP:  10036
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/566,398
;       FILING DATE:  01-DEC-1995
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Frommer, William S.
;       REGISTRATION NUMBER:  25,506
;       REFERENCE/DOCKET NUMBER:  454310-2880
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 840-3333
;       TELEFAX:  (212) 840-0712
;  INFORMATION FOR SEQ ID NO:  39:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  7351 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-566-398-39

  Query Match             11.8%;  Score 28.2;  DB 2;  Length 7351;
  Best Local Similarity   48.4%;  Pred. No. 20;
  Matches   78;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy      28 AAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATACTCCGAA 87
             | ||   ||  || |     | |  |  | ||| || |  |||| | |   ||| ||
Db    6654 ATAAAGTTTAAATTTTTATCATACCAGGTGGAAACAATCTTTCATCGTTATATTCTTTT 6595

Qy      88 ACCACTGGACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATA 147
                | | ||   |||| || |||  || ||  |      | |    |     |    | | |
Db    6594 TTAAGTAGATATATTTGTAAGATAATTATTGAATTTACATATTTATTATCAAAATTAGTC 6535

Qy     148 CTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAAT 188
             || ||  || | || ||||| ||  || ||   |    || |
Db    6534 ATATAAGTAAGAGGAGTATTTCCGTAAGAGTCAGTAATATT 6494


RESULT 15
US-08-458-356-127/c
; Sequence 127, Application US/08458356
; Patent No. 5942235
;  GENERAL INFORMATION:
;     APPLICANT:  Paoletti, Enzo
;     APPLICANT:  Tartaglia, James
;     APPLICANT:  Cox, William I.
;     TITLE OF INVENTION:  RECOMBINANT VIRUS IMMUNOTHERAPY
;     NUMBER OF SEQUENCES:  217
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Curtis, Morris & Safford
;       STREET:  530 Fifth Avenue
;       CITY:  New York
;       STATE:  NY
;       COUNTRY:  USA
;       ZIP:  10036
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/458,356
;       FILING DATE:  02-JUN-1995
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/184,009
;       FILING DATE:  19-JAN-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Frommer, William S.
;       REGISTRATION NUMBER:  25,506
;       REFERENCE/DOCKET NUMBER:  454310-2530
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 840-3333
;       TELEFAX:  (212) 840-0712
;       TELEX:  425066CURTMS
;  INFORMATION FOR SEQ ID NO:  127:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  7351 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-458-356-127

  Query Match             11.8%;  Score 28.2;  DB 2;  Length 7351;
  Best Local Similarity   48.4%;  Pred. No. 20;
  Matches   78;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy      28 AAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATACTCCGAA 87
             | ||   ||  || |     | |  |  | ||| || |  |||| | |   ||| ||
Db    6654 ATAAAGTTTAAATTTTTATCATACCAGGTGGAAACAATCTTTCATCGTTATATTCTTTT 6595

Qy      88 ACCACTGGACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATA 147
                | | ||   |||| || |||  || ||  |      | |    |     |    | | |
Db    6594 TTAAGTAGATATATTTGTAAGATAATTATTGAATTTACATATTTATTATCAAAATTAGTC 6535

Qy     148 CTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAAT 188
             || ||  || | || ||||| ||  || ||   |    || |
Db    6534 ATATAAGTAAGAGGAGTATTTCCGTAAGAGTCAGTAATATT 6494
```

Search completed: May 12, 2003, 02:38:28
Job time : 164.429 secs

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 02:35:34 ; Search time 224.571 Seconds
(without alignments)
1328.082 Million cell updates/sec

Title: US-09-980-277-1
Perfect score: 240
Sequence: 1 gatctgaatccgaccaaccg..........cgcaattatattttgttctt 240

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	35.2	14.7	2000	9	US-09-938-842A-4759	Sequence 4759, Ap
	2	33.2	13.8	10949	9	US-09-764-891-9719	Sequence 9719, Ap
	3	33.2	13.8	10951	9	US-09-764-891-9718	Sequence 9718, Ap
c	4	31.6	13.2	2303	10	US-09-925-302-113	Sequence 113, App
	5	31.4	13.1	4956	9	US-09-815-923-9	Sequence 9, Appli
c	6	31.2	13.0	2000	9	US-09-938-842A-4228	Sequence 4228, Ap
	7	31.2	13.0	640681	10	US-09-790-988-1	Sequence 1, Appli
c	8	31	12.9	527	9	US-10-184-644-522	Sequence 522, App
c	9	31	12.9	527	9	US-10-184-634-522	Sequence 522, App
	10	30.8	12.8	6931	10	US-09-764-877-2137	Sequence 2137, Ap
	11	30.8	12.8	15287	10	US-09-764-877-2134	Sequence 2134, Ap
c	12	30.8	12.8	640681	10	US-09-790-988-1	Sequence 1, Appli
	13	30	12.5	602	10	US-09-998-598-363	Sequence 363, App
c	14	30	12.5	29544	10	US-09-464-767-1	Sequence 1, Appli
c	15	30	12.5	32745	10	US-09-464-767-3	Sequence 3, Appli
c	16	29.8	12.4	8188	10	US-09-070-927A-21	Sequence 21, Appl
	17	29.6	12.3	401	9	US-09-946-807-128	Sequence 128, App
	18	29.6	12.3	401	10	US-09-795-668-128	Sequence 128, App
	19	29.6	12.3	401	10	US-09-795-686-128	Sequence 128, App
c	20	29.6	12.3	458	10	US-09-867-701-6881	Sequence 6881, Ap
c	21	29.6	12.3	513509	9	US-09-754-853A-4	Sequence 4, Appli
	22	29.6	12.3	1503841	9	US-09-946-807-1	Sequence 1, Appli
	23	29.6	12.3	1503841	10	US-09-795-668-1	Sequence 1, Appli
	24	29.6	12.3	1503841	10	US-09-795-686-1	Sequence 1, Appli
c	25	29.4	12.2	184	10	US-09-864-761-33270	Sequence 33270, A
c	26	29.4	12.2	580	10	US-09-864-761-16804	Sequence 16804, A
c	27	29.4	12.2	797	9	US-10-002-344A-50	Sequence 50, Appl
	28	29.4	12.2	1387	9	US-10-067-443-21	Sequence 21, Appl
c	29	29.4	12.2	1419	10	US-09-925-300-311	Sequence 311, App
c	30	29.4	12.2	1458	9	US-09-938-842A-3615	Sequence 3615, Ap
c	31	29.4	12.2	2000	9	US-09-938-842A-4439	Sequence 4439, Ap
	32	29.4	12.2	14364	9	US-10-067-443-20	Sequence 20, Appl
	33	29.2	12.2	503	10	US-09-867-701-6750	Sequence 6750, Ap
	34	29.2	12.2	1791	10	US-09-974-300-2178	Sequence 2178, Ap
c	35	29.2	12.2	4374	10	US-09-772-316-5	Sequence 5, Appli
	36	29	12.1	1812	10	US-09-728-711-10	Sequence 10, Appl
c	37	29	12.1	3775	7	US-08-781-986A-238	Sequence 238, App
c	38	28.8	12.0	2000	9	US-09-938-842A-3854	Sequence 3854, Ap
	39	28.6	11.9	451	9	US-09-918-995-9958	Sequence 9958, Ap
c	40	28.6	11.9	1271	10	US-09-037-460-1	Sequence 1, Appli
c	41	28.6	11.9	2000	9	US-09-938-842A-5041	Sequence 5041, Ap
	42	28.6	11.9	174424	10	US-09-967-768A-314	Sequence 314, App
	43	28.6	11.9	203654	10	US-09-820-905-3	Sequence 3, Appli
c	44	28.4	11.8	279	10	US-09-878-574-7375	Sequence 7375, Ap
	45	28.4	11.8	396	10	US-09-983-965-4115	Sequence 4115, Ap

ALIGNMENTS

RESULT 1
US-09-938-842A-4759
; Sequence 4759, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAIN
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4759
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4759

```
  Query Match             14.7%;  Score 35.2;  DB 9;  Length 2000;
  Best Local Similarity   58.7%;  Pred. No. 1.2;
  Matches   61;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy    136 TGCCTAAAAATACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAAC 195
          ||  || |||||    |||||     || ||||   | | |||      || ||| || |
Db    555 TGGTTATAAATATATGAAACATTTACATAATTTAGTAAAACGTTCAACTTGATAAAATAA 614

Qy    196 AAGAACTAATAAACGCTCTCGCACCCGCAATTATATTTTGTTCT 239
          |  ||||||  ||  |||| | |    |||  | | |||| ||||
Db    615 ATTAACTAAATAAACCTCTTGTATAGGCATCTGTCTTTTATTCT 658
```

RESULT 2
US-09-764-891-9719

```
; Sequence 9719, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC006
;  CURRENT APPLICATION NUMBER: US/09/764,891
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 10231
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9719
;   LENGTH: 10949
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-891-9719

  Query Match            13.8%;  Score 33.2;  DB 9;  Length 10949;
  Best Local Similarity  48.9%;  Pred. No. 10;
  Matches   89;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy       11 CGACCAACCGTTCTATGAAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTT 70
             | | | || | |  ||| ||    ||| || |      |||    |  ||   |      ||
Db     9927 CAAACTACTGATACATGTAACAAATTGCATGGACCTCAAGGGAATTATGCTTGGGTGGTT 9986

Qy       71 TAGCAACATACTCCGAAACCACTGGACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGG 130
              | |    | | ||| |||     |  ||      | | ||  ||||| | | | |   |||
Db     9987 TGGGGGGACAATCCTAAAGTTTTACACACTGTATGATTCTGTTTTTGTAACATTCTTAAA 10046

Qy      131 TGACCTGCCTAAAAATACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAATAT 190
              | |      ||| | | |      |||||| ||     |  |||| ||  | ||||  | ||
Db    10047 TAATAAAGCTATAGAGATGGAAACAAGATTCGTGGTTGCCAGGTCTTGGGACTGAAGGG 10106

Qy      191 AA 192
            ||
Db    10107 AA 10108

RESULT 3
US-09-764-891-9718
; Sequence 9718, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC006
;  CURRENT APPLICATION NUMBER: US/09/764,891
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 10231
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9718
;   LENGTH: 10951
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-891-9718

  Query Match            13.8%;  Score 33.2;  DB 9;  Length 10951;
  Best Local Similarity  48.9%;  Pred. No. 10;
  Matches   89;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy       11 CGACCAACCGTTCTATGAAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTT 70
             | | | || | |  ||| ||    ||| || |      |||    |  ||   |      ||
Db     9929 CAAACTACTGATACATGTAACAAATTGCATGGACCTCAAGGGAATTATGCTTGGGTGGTT 9988

Qy       71 TAGCAACATACTCCGAAACCACTGGACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGG 130
              | |    | | ||| |||     |  ||      | | ||  ||||| | | | |   |||
Db     9989 TGGGGGGACAATCCTAAAGTTTTACACACTGTATGATTCTGTTTTTGTAACATTCTTAAA 10048

Qy      131 TGACCTGCCTAAAAATACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAATAT 190
              | |      ||| | | |      |||||| ||     |  |||| ||  | ||||  | ||
Db    10049 TAATAAAGCTATAGAGATGGAAACAAGATTCGTGGTTGCCAGGTCTTGGGACTGAAGGG 10108

Qy      191 AA 192
            ||
Db    10109 AA 10110

RESULT 4
US-09-925-302-113/c
; Sequence 113, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA104
;  CURRENT APPLICATION NUMBER: US/09/925,302
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05918
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 896
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 113
;   LENGTH: 2303
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-302-113

  Query Match            13.2%;  Score 31.6;  DB 10;  Length 2303;
  Best Local Similarity  53.2%;  Pred. No. 15;
  Matches   67;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy       25 ATGAAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATACTCC 84
            || || ||| |  |           | |    |||  |  ||    || || ||   || || |
Db     1105 ATAAAGATCATAATTGGCCTATTATCCCTGTATTTCATATTATTTCTAAAAATATTCCTA 1046

Qy       85 GAAACCACTGGACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAA 144
            ||||     ||| | | | ||  ||||  |   |  ||| ||   ||| ||     ||   || ||||
Db     1045 GAAATATCTGCAAGAATAGTTATTTTAATCTTTTAAGTAGCTGTGAAAACAAACCAAAAA 986

Qy      145 ATACTA 150
            ||  ||
Db      985 ATTGTA 980

RESULT 5
US-09-815-923-9
; Sequence 9, Application US/09815923
; Publication No. US20020197644A1
; GENERAL INFORMATION:
;  APPLICANT: Gill, Sarjeet S.
;  APPLICANT:  Ross, Linda S.
;  APPLICANT:  The Regents of the University of California
;  TITLE OF INVENTION: Use of Insect Cell Membrane Transporters as No. US2002019764
;  TITLE OF INVENTION:  Target Sites for Insecticides
;  FILE REFERENCE: 023070-093800US
;  CURRENT APPLICATION NUMBER: US/09/815,923
;  CURRENT FILING DATE:  2001-03-23
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
;   LENGTH: 4956
;   TYPE: DNA
;   ORGANISM: Manduca sexta
;   FEATURE:
;   OTHER INFORMATION: neurotransmitter transporter encoded by inebriated
;   OTHER INFORMATION: gene
US-09-815-923-9

  Query Match            13.1%;  Score 31.4;  DB 9;  Length 4956;
  Best Local Similarity  54.9%;  Pred. No. 25;
  Matches   62;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy       48 TAGGCTGGTTGGCAAGATACTTTTAGCAACATACTCCGAAACCACTGGACGGATTTTTAT 107
```

```
                ||||||||   |||  || |  || |  | || |     || ||| | |   ||| || |
Db    3488 TAGGCTGGTCTACAACCTAGTAATACCTATGTAGTGTTGAAACACCGAATAAATTATTTT 3547

Qy     108 GATGTTTGTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAACAAAAT 160
              || ||||   |     | | |    | | || | ||| | | |||| || |
Db    3548 AATCGTTGTGGAAATTTGTGTAGATCCTTACCGATAAAGATTTCAAAATAACT 3600


RESULT 6
US-09-938-842A-4228/c
; Sequence 4228, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
;  APPLICANT: Harper, Jeff
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Wang, Xun
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
;  TITLE OF INVENTION:  SAME, AND METHODS OF USE
;  FILE REFERENCE: SCRIP1300-3
;  CURRENT APPLICATION NUMBER: US/09/938,842A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 60/227,866
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: US 60/264,647
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/300,111
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4228
;   LENGTH: 2000
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-938-842A-4228

  Query Match             13.0%;  Score 31.2;  DB 9;  Length 2000;
  Best Local Similarity   63.2%;  Pred. No. 19;
  Matches   48;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy      97 CGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAACA 156
              | || | || ||  || ||||||| | |||     |||  |    ||| | || |    || |
Db     661 CAGACTATTTTGGAGTCTGTTTTATATGCTGGTTTGAATTCGATAATATTAATGATAATA 602

Qy     157 AAATGATTATTTGCGA 172
           |||| || | ||||||
Db     601 AAATCATCAATTGCGA 586


RESULT 7
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
;  APPLICANT: SHIGENOBU, SHUJI
;  APPLICANT:  WATANABE, HIDEMI
;  APPLICANT:  HATTORI, MASAHIRA
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;  FILE REFERENCE: 081356/0159
;  CURRENT APPLICATION NUMBER: US/09/790,988
;  CURRENT FILING DATE:  2001-02-23
;  PRIOR APPLICATION NUMBER: JP2000-107160
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

  Query Match             13.0%;  Score 31.2;  DB 10;  Length 640681;
  Best Local Similarity   48.8%;  Pred. No. 2.6e+02;
```

```
  Matches   84;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy      35 TTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATACTCCGAAACCACTG 94
             || ||  | |  |    || |      || |  |||||| |     | | |    ||     |
Db   83625 TTTAATAGAAATATTACCTTATAAAAAAAAATCTTTTATAATTTTTCCCTATAATTTTTT 83684

Qy      95 GACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAA 154
             | |  || ||    ||| |||  ||| |||||   ||||| |        |  |||||| |||
Db   83685 GTCTAATATTATAGATATTTTATTTGGAGGCCAGGGTGATTTTAAAGATCATACTAAAAA 83744

Qy     155 CAAAATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAACTAATA 206
            | ||    ||||      | | |     |   ||  ||| ||  | | | |
Db   83745 AATAAGAGATATTACTTCTACTGAGTCTTTAGTTAATAAAAAAATAATGAAA 83796


RESULT 8
US-10-184-644-522/c
; Sequence 522, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C227
;  CURRENT APPLICATION NUMBER: US/10/184,644
;  CURRENT FILING DATE:  2002-06-28
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 522
;   LENGTH: 527
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-184-644-522

  Query Match             12.9%;  Score 31;  DB 9;  Length 527;
  Best Local Similarity   10.3%;  Pred. No. 12;
  Matches   21;  Conservative   72;  Mismatches  110;  Indels    0;  Gaps    0;

Qy      38 TATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATACTCCGAAACCACTGGAC 97
              :|    :   :    |         |   : |     ::  ::   :   :: |:: :: |:::::
Db     521 YAMN.M.GS...GMA....TATBGA...C.BH.SRD..WA.HDTTSY.DMTCMDYKB... 462

Qy      98 GGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAACAA 157
                |::  ::  ::: :: ::|  :::::  :  :     :  |:    :| |: : :
Db     461 W.BTYH..MB..HDD.YS.YKTN.MRSSHA.BAY.TY..H.SAKAM.N..BTTCMTMKDT 402

Qy     158 AATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAACTAATAAACGCTCTCGC 217
              :  :     : :: : :  | : | :  :     :|: :   :     ::: :  || :
Db     401 .NH..HC...BCBK.BCDR.T.R.CNKCCDA..TMAM.DC..HN...WHR.YANTC.AS. 342

Qy     218 ACCCGCAATTATATTTTGTTCTT 240
             ::|:::: :      | |:|   |
Db     341 MMCMRYW.BM....T.TST..NT 319


RESULT 9
US-10-184-634-522/c
; Sequence 522, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
```

```
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C217
;  CURRENT APPLICATION NUMBER: US/10/184,634
;  CURRENT FILING DATE:  2002-06-28
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 522
;   LENGTH: 527
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-184-634-522

  Query Match             12.9%;  Score 31;  DB 9;  Length 527;
  Best Local Similarity   10.3%;  Pred. No. 12;
  Matches   21;  Conservative   72;  Mismatches  110;  Indels    0;  Gaps    0;

Qy       38 TATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATACTCCGAAACCACTGGAC 97
            :|  :  :   |     | : |     ::  ::  :  :: |:: :: |:::::
Db      521 YAMN.M.GS...GMA....TATBGA...C.BH.SRD..WA.HDTTSY.DMTCMDYKB... 462

Qy       98 GGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAACAA 157
              |::  ::  ::: :: ::|  :::::  : :  :      :  |:    :| |: : :
Db      461 W.BTYH..MB..HDD.YS.YKTN.MRSSHA.BAY.TY..H.SAKAM.N..BTTCMTMKDT 402

Qy      158 AATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAACTAATAAACGCTCTCGC 217
              :   :    : :: : :  | : | :  :     :|: :    :     ::: :  || :
Db      401 .NH..HC...BCBK.BCDR.T.R.CNKCCDA..TMAM.DC..HN...WHR.YANTC.AS. 342

Qy      218 ACCCGCAATTATATTTTGTTCTT 240
            ::|::::  :      | |:|    |
Db      341 MMCMRYW.BM....T.TST..NT 319


RESULT 10
US-09-764-877-2137
; Sequence 2137, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC005
;  CURRENT APPLICATION NUMBER: US/09/764,877
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 4031
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2137
;   LENGTH: 6931
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-877-2137

  Query Match             12.8%;  Score 30.8;  DB 10;  Length 6931;
  Best Local Similarity   50.7%;  Pred. No. 45;
  Matches   74;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy       20 GTTCTATGAAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACAT 79
              || || | |||  ||  |  |  ||| | |  |  |||  || |||| |     | |
Db     2843 GTGTTAAGTAAAAGATTAAAAATGCACTGATGAAGCATGGGTTGAAACTTCTCAATATAG 2902

Qy       80 ACTCCGAAACCACTGGACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCC 139
               ||  |  |||    | |   ||| ||||||  ||     | | ||  | |  | |
Db     2903 TCTGAGGCACCTTAGCAACAATTATTATGAAATTGTCATCAAAGCGTGTCTTCATGCTTA 2962

Qy      140 TAAAAATACTACAAACAAAATGATTA 165
             ||    || || || ||| | ||| ||
Db     2963 TAGTCATTCTTCACACATATTGAGTA 2988


RESULT 11
US-09-764-877-2134
; Sequence 2134, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC005
;  CURRENT APPLICATION NUMBER: US/09/764,877
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 4031
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2134
;   LENGTH: 15287
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-877-2134

  Query Match             12.8%;  Score 30.8;  DB 10;  Length 15287;
  Best Local Similarity   50.7%;  Pred. No. 65;
  Matches   74;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy       20 GTTCTATGAAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACAT 79
              || || | |||  ||  |  |  ||| | |  |  |||  || |||| |     | |
Db    11199 GTGTTAAGTAAAAGATTAAAAATGCACTGATGAAGCATGGGTTGAAACTTCTCAATATAG 11258

Qy       80 ACTCCGAAACCACTGGACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCC 139
               ||  |  |||    | |   ||| ||||||  ||     | | ||  | |  | |
Db    11259 TCTGAGGCACCTTAGCAACAATTATTATGAAATTGTCATCAAAGCGTGTCTTCATGCTTA 11318

Qy      140 TAAAAATACTACAAACAAAATGATTA 165
             ||    || || || ||| | ||| ||
Db    11319 TAGTCATTCTTCACACATATTGAGTA 11344


RESULT 12
US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
;  APPLICANT: SHIGENOBU, SHUJI
;  APPLICANT:  WATANABE, HIDEMI
;  APPLICANT:  HATTORI, MASAHIRA
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;  FILE REFERENCE: 081356/0159
;  CURRENT APPLICATION NUMBER: US/09/790,988
;  CURRENT FILING DATE:  2001-02-23
;  PRIOR APPLICATION NUMBER: JP2000-107160
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

  Query Match             12.8%;  Score 30.8;  DB 10;  Length 640681;
  Best Local Similarity   53.3%;  Pred. No. 3.3e+02;
  Matches   65;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy       46 CTTAGGCTGGTTGGCAAGATACTTTTAGCAACATACTCCGAAACCACTGGACGGATTTTT 105
            ||||    | ||     || ||  |||||| || |||        |  ||||  | |||| | |
Db   609177 CTTATCAAGATTTAAAAAATCATTTTAGAAAAATAAAAAATATTCACTTAAGGGATCTAT 609118

Qy      106 ATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAACAAAATGATTA 165
```

```
            ||      |  |||  |     |  |||       |   |    |  |  ||  |    |    |||    |||||  |  |
Db 609117 TTGCATCTGATTTGAATCGATTTAAAAAATTTTCAATAATTTTGAAAATGAAATGTTAA 609058

Qy    166 TT 167
          ||
Db 609057 TT 609056


RESULT 13
US-09-998-598-363
; Sequence 363, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
;  APPLICANT: Stolk, John A.
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Chenault, Ruth A.
;  APPLICANT:  Meagher, Madelein Joy
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF COLON CANCER
;  FILE REFERENCE: 210121.561
;  CURRENT APPLICATION NUMBER: US/09/998,598
;  CURRENT FILING DATE:  2001-11-16
;  NUMBER OF SEQ ID NOS: 2606
;  SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 363
;   LENGTH: 602
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-998-598-363

  Query Match             12.5%;  Score 30;  DB 10;  Length 602;
  Best Local Similarity   54.8%;  Pred. No. 25;
  Matches   57;  Conservative    1;  Mismatches   46;  Indels    0;  Gaps    0;

Qy      97 CGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAACA 156
            || |   |||    |   |||:| |  |    | |||  | ||||       | |||  |    | |||  |
Db     499 CGCACCATTAACTTTGTTGKTGTTGKTGKTTTTGAGACCCTGTCACAAAAAAAAAAAAAGA 558

Qy     157 AAATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAA 200
            |  |  ||||  |        ||         |||   ||||  |  ||||  ||  ||
Db     559 AGAAGATTCTACTTTAGCTTCCGGGATTTAAAAAAAAAAAAAAA 602


RESULT 14
US-09-464-767-1/c
; Sequence 1, Application US/09464767
; Patent No. US20020045249A1
; GENERAL INFORMATION:
;  APPLICANT: Both, Gerald
;  APPLICANT:  Boyle, David
;  APPLICANT:  Vrati, Sudhanshu
;  TITLE OF INVENTION: DNA Encoding Ovine Adenovirus (OAV287) and Its Use as a Viral Vec
;  FILE REFERENCE: 50179-073
;  CURRENT APPLICATION NUMBER: US/09/464,767
;  CURRENT FILING DATE:  1999-12-16
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 29544
;   TYPE: DNA
;   ORGANISM: Ovine adenovirus
US-09-464-767-1

  Query Match             12.5%;  Score 30;  DB 10;  Length 29544;
  Best Local Similarity   59.3%;  Pred. No. 1.5e+02;
  Matches   51;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy      34 GTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATACTCCGAAACCACT 93
            ||| ||||||    |||   |||| | | ||| || |    |     || | | || |||||
Db   12259 GTTTTATCTACCATTAATCTGGCTAGAAAGCTATTACTGTTAAAAAATTCTGAAACATTG 12200

Qy      94 GGACGGATTTTTATGATGTTTGTTTT 119
            || |    || || | ||     |||||
Db   12199 GGGCAATTTGTTTTTATAAAAGTTTT 12174


RESULT 15
US-09-464-767-3/c
; Sequence 3, Application US/09464767
; Patent No. US20020045249A1
; GENERAL INFORMATION:
;  APPLICANT: Both, Gerald
;  APPLICANT:  Boyle, David
;  APPLICANT:  Vrati, Sudhanshu
;  TITLE OF INVENTION: DNA Encoding Ovine Adenovirus (OAV287) and Its Use as a Vir
;  FILE REFERENCE: 50179-073
;  CURRENT APPLICATION NUMBER: US/09/464,767
;  CURRENT FILING DATE:  1999-12-16
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
;   LENGTH: 32745
;   TYPE: DNA
;   ORGANISM: synthetic construct
US-09-464-767-3

  Query Match             12.5%;  Score 30;  DB 10;  Length 32745;
  Best Local Similarity   59.3%;  Pred. No. 1.6e+02;
  Matches   51;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy      34 GTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATACTCCGAAACCACT 93
            ||| ||||||    |||   |||| | | ||| || |    |     || | | || |||||
Db   12258 GTTTTATCTACCATTAATCTGGCTAGAAAGCTATTACTGTTAAAAAATTCTGAAACATTG 12199

Qy      94 GGACGGATTTTTATGATGTTTGTTTT 119
            || |    || || | ||     |||||
Db   12198 GGGCAATTTGTTTTTATAAAAGTTTT 12173


Search completed: May 12, 2003, 04:47:58
Job time : 530.571 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 01:04:28 ; Search time 2929.71 Seconds
(without alignments)
1326.722 Million cell updates/sec

Title: US-09-980-277-1
Perfect score: 240
Sequence: 1 gatctgaatccgaccaaccg.....cgcaattatattttgttctt 240

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44.2	18.4	687	17	BH210086	BH210086 Sm1-56J20
2	40.8	17.0	600	17	AQ343444	AQ343444 RPCI11-12
c 3	36.8	15.3	607	13	BI351574	BI351574 IP1_72_F0
4	36.6	15.2	688	17	AZ370732	AZ370732 1M0122E12
5	36.6	15.2	1099	17	CNS00167	AL074301 Drosophil
6	36.4	15.2	601	13	BM161288	BM161288 EST563811
c 7	36.4	15.2	696	13	BJ439775	BJ439775 BJ439775
c 8	35.8	14.9	483	13	BJ437738	BJ437738 BJ437738
c 9	35.8	14.9	655	13	BJ383574	BJ383574 BJ383574
10	35.6	14.8	661	17	AZ526385	AZ526385 254Pdb05
c 11	35.4	14.8	524	9	AL728080	AL728080 AL728080
c 12	35.4	14.8	772	13	BJ356150	BJ356150 BJ356150
c 13	35.2	14.7	541	13	BJ444759	BJ444759 BJ444759
c 14	35.2	14.7	549	13	BJ444864	BJ444864 BJ444864
c 15	35.2	14.7	594	13	BJ440866	BJ440866 BJ440866
c 16	35.2	14.7	606	13	BJ357820	BJ357820 BJ357820
c 17	35.2	14.7	647	13	BJ347241	BJ347241 BJ347241
c 18	35.2	14.7	653	13	BJ355861	BJ355861 BJ355861
c 19	35.2	14.7	653	13	BJ444955	BJ444955 BJ444955
c 20	35.2	14.7	657	13	BJ442909	BJ442909 BJ442909
c 21	35.2	14.7	662	13	BJ438327	BJ438327 BJ438327
c 22	35.2	14.7	684	13	BJ340803	BJ340803 BJ340803
c 23	35.2	14.7	687	13	BJ443059	BJ443059 BJ443059
c 24	35.2	14.7	698	13	BJ441393	BJ441393 BJ441393
c 25	35.2	14.7	726	13	BJ379843	BJ379843 BJ379843
c 26	35.2	14.7	726	13	BJ437903	BJ437903 BJ437903
c 27	35.2	14.7	790	13	BJ438503	BJ438503 BJ438503
c 28	35.2	14.7	792	13	BJ350365	BJ350365 BJ350365
c 29	35.2	14.7	800	13	BJ436508	BJ436508 BJ436508
30	35.2	14.7	800	14	CB2262	CB2262 CB2262 Leuk
31	35.2	14.7	840	17	AZ531356	AZ531356 EMTCV49TF
32	35	14.6	116	9	AI966392	AI966392 ac38d03.y
33	35	14.6	523	17	AQ506546	AQ506546 RPCI-11-2
34	35	14.6	576	17	AQ542264	AQ542264 RPCI-11-3
c 35	35	14.6	695	17	CNS03DRP	AL139470 Tetraodon
36	35	14.6	993	17	AQ750012	AQ750012 HS_5574_A
37	35	14.6	1101	17	CNS0578R	AL324324 Tetraodon
38	34.8	14.5	474	17	AQ546802	AQ546802 CITRI-E1-
c 39	34.8	14.5	609	13	BJ442768	BJ442768 BJ442768
40	34.8	14.5	1099	17	CNS05P3Y	AL347479 Tetraodon
c 41	34.4	14.3	468	14	BQ522921	BQ522921 NISC_n119
c 42	34.4	14.3	594	13	BJ357131	BJ357131 BJ357131
c 43	34.4	14.3	607	13	BJ440342	BJ440342 BJ440342
c 44	34.4	14.3	742	13	BJ349413	BJ349413 BJ349413
c 45	34.4	14.3	753	13	BJ441379	BJ441379 BJ441379

ALIGNMENTS

RESULT 1
BH210086
LOCUS BH210086 687 bp DNA linear GSS 24-OCT-2001
DEFINITION Sm1-56J20.TF Sm1 Schistosoma mansoni genomic clone Sm1-56J20, DNA sequence.
ACCESSION BH210086
VERSION BH210086.1 GI:16389299
KEYWORDS GSS.
SOURCE Schistosoma mansoni.
ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE 1 (bases 1 to 687)
AUTHORS Shetty,J., Simpson,A., Malek,J., Koo,H., LoVerde,P.T. and El-Sayed,N.M.
TITLE Use of end sequences from Schistosoma mansoni (Puerto Rico strain) Sm1 BAC library for gene discovery and map construction
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: Sm1-56J20.TR
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Ic.edu).
Seq primer: M13 For
Class: BAC ends.

```
FEATURES             Location/Qualifiers
     source          1..687
                     /organism="Schistosoma mansoni"
                     /strain="Puerto Rico"
                     /db_xref="taxon:6183"
                     /clone="Sm1-56J20"
                     /clone_lib="Sm1"
                     /note="Vector: pBeloBAC11; Site_1: Hin dIII; Constructed
                     in the laboratory of Dr. Denis Le Paslier at the Fondation
                     Jean Dausset, CEPH, Paris, France. Briefly, Schistosoma
                     mansoni agarose embedded DNA was partially digested with
                     Hin dIII. High molecular weight fragments were ligated in
                     pBeloBAC11 digested with Hin dIII. The average insert size
                     is 100 kb. Total clone coverage: approx. 7.95 X the
                     haploid genome. Further information can be found in Le
                     Paslier et al. (2000) Construction and characterization of
                     a Schistosoma mansoni bacterial artificial chromosome
                     library. Genomics 65: 87-94."
BASE COUNT     218 a    101 c    152 g    216 t
ORIGIN

  Query Match              18.4%;  Score 44.2;  DB 17;  Length 687;
  Best Local Similarity    93.9%;  Pred. No. 0.39;
  Matches   46;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        81 CTCCGAAACCACTGGACGGATTTTTATGATGTTTGTTTTAGAGGCTTTG 129
             ||||||||||||||||||||||||||||||||||||||||||   ||||
Db       388 CTCCGAAACCACTGGACGGATTTTTATGATGTTTGTTTTAGATTATTTG 436


RESULT 2
AQ343444
LOCUS       AQ343444                 600 bp    DNA     linear   GSS 07-MAY-1999
DEFINITION  RPCI11-124D13.TV RPCI-11 Homo sapiens genomic clone RPCI-11-124D13,
            DNA sequence.
ACCESSION   AQ343444
VERSION     AQ343444.1  GI:4168340
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI11-124D13.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..600
                     /organism="Homo sapiens"
                     /db_xref="GDB:7547316"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-124D13"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT     257 a     77 c     95 g    171 t
ORIGIN

  Query Match              17.0%;  Score 40.8;  DB 17;  Length 600;
  Best Local Similarity    49.5%;  Pred. No. 2.8;
  Matches  105;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy        22 TCTATGAAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATAC 81
              | ||| | | || |  ||     || |||    | || |   | | || |   |  ||| ||
Db        45 TATATTACACTCTTAATAAGATGATTTAATATTGTAGAAATGTTAATCATTTCAAATTAA 104

Qy        82 TCCGAAACCACTGGACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTA 141
              ||   |  || |  |      |  |  |  ||  |||  ||  |        |  ||  ||    |
Db       105 TCTATATTCAGTTCAATCCTAATAAAAATTATTGAATTTAATAAACTTATGCACTTAAAA 164

Qy       142 AAAATACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAAC 201
                ||||| || |   ||| |  ||| |       |      |    | |||||| | ||  ||||
Db       165 TGTATACTGCAGAATAAAGGTCTATGTATAGGTTAATCAACCTTAATAAAGAAGTAGAAA 224

Qy       202 TAATAAACGCTCTCGCACCCGCAATTATATTT 233
               ||||||   | | |      | || | |||| |
Db       225 AGATAAACTAACCCTCTAACACACTAATATAT 256


RESULT 3
BI351574/c
LOCUS       BI351574                 607 bp    mRNA    linear   EST 31-JUL-2001
DEFINITION  IP1_72_F04.b1_A002 Immature pannicle 1 (IP1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   BI351574
VERSION     BI351574.1  GI:15046016
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 607)
  AUTHORS   Klein,R.R., Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: developing preanthesis pannicles
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for high quality sequence is
            20. Three-prime sequences, which are obtained with PolyTMix or T7
            sequencing primer, are presented as the reverse complement.
            Seq primer: JEN REV
            High quality sequence stop: 599
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..607
                     /organism="Sorghum bicolor"
                     /cultivar="BTx623"
                     /db_xref="taxon:4558"
                     /clone_lib="Immature pannicle 1 (IP1)"
                     /note="Organ: Developing preanthesis pannicles; Vector:
                     pBluescript II SK(-) from Lambda Zap II; Site_1: XhoI;
                     Site_2: EcoRI; The library was made from poly-A RNA in the
                     cloning vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT     142 a    159 c    142 g    164 t
ORIGIN

  Query Match              15.3%;  Score 36.8;  DB 13;  Length 607;
```

```
  Best Local Similarity   54.4%;  Pred. No. 28;
  Matches   74;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy       28 AAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATACTCCGAA 87
            |||||  |||||| | | ||||||  | ||  |   |   |||  || |      ||  |
Db      521 AAAATTCTTGTATGTGGGCTTAGGGCGCTTTACCTTACCATTTGGGCCATTACGACCAGA 462

Qy       88 ACCACTGGACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATA 147
            ||||   |   |   | | | | | ||| ||   |   |||||| ||| |  ||||  |
Db      461 ACCAAGAGGACGGCCTCTCTTACGCTTGGTTACCAAAGTTTGGTTACCCGAGTAAAGCTC 402

Qy      148 CTACAAACAAAATGAT 163
             |    || || ||||
Db      401 ATCATCACCAAGTGAT 386


RESULT 4
AZ370732
LOCUS       AZ370732                 688 bp    DNA     linear   GSS 02-OCT-2000
DEFINITION  1M0122E12F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0122E12 F, DNA sequence.
ACCESSION   AZ370732
VERSION     AZ370732.1  GI:10484528
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 688)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL . Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0122  row: E  column: 12
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 688.
FEATURES             Location/Qualifiers
     source          1..688
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0122E12"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      249 a    128 c    119 g    192 t
ORIGIN

  Query Match             15.2%;  Score 36.6;  DB 17;  Length 688;
  Best Local Similarity   53.1%;  Pred. No. 31;
  Matches   78;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy       87 AACCACTGGACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAAT 146
            | |||| ||   | ||  | |  || || ||| | |||||    ||| |  | |||||
Db      331 AGCCACAGGGATGCTTGCTTTAGTGGTTTTTTAAAAGGCTACTTTGAACAACATAAAATC 390

Qy      147 ACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAACTAATA 206
            | | |||| | || |    ||||       ||      | || | |||  | || || |
Db      391 AATTCAAATAGAAAGCCACTTTGAACCCAAGTCTCTACTCATGTTAAAAGAAAAATATTC 450

Qy      207 AACGCTCTCGCACCCGCAATTATATTT 233
            | |    |  || |  |||| ||  ||
Db      451 AGCCAGATGACAACGCCAATGATTATT 477


RESULT 5
CNS00I6T
LOCUS       CNS00I6T                1099 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR36B11 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL074301
VERSION     AL074301.1  GI:4954182
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1099)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1..1099
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR36B11"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
BASE COUNT      330 a    150 c    135 g    327 t    157 others
ORIGIN

  Query Match             15.2%;  Score 36.6;  DB 17;  Length 1099;
  Best Local Similarity   40.7%;  Pred. No. 28;
  Matches   55;  Conservative   26;  Mismatches   54;  Indels    0;  Gaps    0;
```

```
Qy      25 ATGAAAATCGTTGTATCTAGACTTAGGCTGGTTGGCAAGATACTTTTAGCAACATACTCC 84
           | ||||||  || | | |       ||      |  |:||  :| :: :|:|: :|  |:   :|
Db     784 AAGAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTKGGRGRGGHRGWTWTMCMACWAKTGCMC 843

Qy      85 GAAACCACTGGACGGATTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAA 144
           | |||||| | |     || |:| |: | ::|:|||: :   | :::||:::   : : |||:
Db     844 GGAACCACCGAATAWATATKTTTKTTTKKTKTTTKKKKTGKKKKGGKBDKKWTYATAAVC 903

Qy     145 ATACTACAAACAAAA 159
              | ||  | ||
Db     904 CACACAAAACGACAA 918
```

```
RESULT 6
BM161288
LOCUS       BM161288                 601 bp    mRNA    linear   EST 04-DEC-2001
DEFINITION  EST563811 PyBS Plasmodium yoelii yoelii cDNA clone PYCKD04 5' end,
            mRNA sequence.
ACCESSION   BM161288
VERSION     BM161288.1  GI:17306969
KEYWORDS    EST.
SOURCE      Plasmodium yoelii yoelii.
  ORGANISM  Plasmodium yoelii yoelii
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 601)
  AUTHORS   Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
            Fraser,C.M. and Carucci,D.J.
  TITLE     Plasmodium yoelii EST project at TIGR
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Jane Carlton
            Parasite Genomics Group
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-530-9319
            Fax: 301-838-0208
            Email: carlton@tigr.org
            For clone info, please contact the Malaria Research and Reference
            Reagent Resource Center, ATCC
            http://www.malaria.mr4.org/mr4pages/index.html
            Seq primer: ADF.
FEATURES             Location/Qualifiers
     source          1..601
                     /organism="Plasmodium yoelii yoelii"
                     /strain="17XL"
                     /db_xref="taxon:73239"
                     /clone="PYCKD04"
                     /clone_lib="PyBS"
                     /dev_stage="Asexual blood stages"
                     /lab_host="E. coli XL-1 Blue"
                     /note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was
                     collected from BALB/cByJ mice infected with Py17XL
                     parasites, and leukocytes removed by passage over
                     microcrystalline cellulose collumns. Total RNA was
                     isolated using the guanidinium isothiocyanate method, and
                     mRNA isolated using oligo(dT)-cellulose chromatography.
                     First strand cDNA synthesis was completed using a 50-base
                     primer and reverse transcriptase in the presence of
                     5-methyl dCTP. After second strand synthesis, uneven
                     termini were treated with Pfu DNA polymerase and EcoRI
                     adaptors ligated to the blunt ends. The sample was cleaved
                     with XhoI and separated on a Sephacryl S-500 column.
                     Size-fractionated cDNA was precipitated and ligated to
                     HybriZAP arms directionally using EcoRI-XhoI cleaved arms.
                     After packaging, the phagemid vector (pAD-GAL4) was
                     excised from the HybriZAP vector and plasmid DNA
                     isolated."
BASE COUNT      292 a     45 c     89 g    175 t
ORIGIN

  Query Match             15.2%;  Score 36.4;  DB 13;  Length 601;
  Best Local Similarity   58.2%;  Pred. No. 36;
  Matches   64;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;
```

```
Qy      99 GATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAACAAA 158
           || |||| ||||  |       |||| || |||| || |      |  ||||       || || |||
Db      35 GAATTTTGTGATAATATATTTATAGACTTTTGTAAAAATCAGAAAATGGTAACGAAAAAA 94

Qy     159 ATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAACTAATAAA 208
           |||| |||| || | |                |||  | |||| || ||  ||  ||
Db      95 ATGAATATTCGCAAAAAAAAAATTAATAAAAAAAAAAAAAAAAAAAAAAGAA 144
```

```
RESULT 7
BJ439775/c
LOCUS       BJ439775                 696 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ439775 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv41f17 3', mRNA sequence.
ACCESSION   BJ439775
VERSION     BJ439775.1  GI:19414497
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 696)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..696
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv41f17"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
                     /sex="mat A"
                     /dev_stage="Growth phase"
BASE COUNT      214 a    113 c     99 g    263 t      7 others
ORIGIN

  Query Match             15.2%;  Score 36.4;  DB 13;  Length 696;
  Best Local Similarity   57.1%;  Pred. No. 34;
  Matches   64;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;
```

```
Qy     125 CTTTGGTGACCTGCCTAAAAATACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGT 184
            | || ||||| |     ||| |||| ||  |     ||| || | | ||    || || |||    |
Db     575 CGTTAGTGACTTTGATAATAATAATAATACTGAAAAGAGTTTTGTTGAAAGAGTGATTGG 516

Qy     185 TAATATAAAACAAGAACTAATAAACGCTCTCGCACCCGCAATTATATTTTGT 236
            |||| |||| |||||     || ||     |  |  |  |    |   ||| ||   | |
Db     515 TAATGTAAATCAAGACNCAACAATTTCATTTGAATATGGTGTTAGATCNTCT 464
```

```
RESULT 8
BJ437738/c
LOCUS       BJ437738                 483 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ437738 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv35a08 3', mRNA sequence.
ACCESSION   BJ437738
VERSION     BJ437738.1  GI:19412460
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 483)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
```

```
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .483
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv35a08"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
                     /sex="mat A"
                     /dev_stage="Growth phase"
BASE COUNT     144 a     79 c     71 g    179 t     10 others
ORIGIN

  Query Match            14.9%;  Score 35.8;  DB 13;  Length 483;
  Best Local Similarity  57.1%;  Pred. No. 53;
  Matches   64;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy    125 CTTTGGTGACCTGCCTAAAAATACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGT 184
          | || ||||| |   ||| |||| ||  |    ||| || | ||   || || |||    |
Db    438 CGTTAGTGACTTTGATAATAATAATAATACTGAAAAGAGTTTTGTTGAAAGAGTGATTGG 379

Qy    185 TAATATAAAACAAGAACTAATAAACGCTCTCGCACCCGCAATTATATTTTGT 236
          |||| |||| |||||    || ||    |  | | |    |    ||| ||   | |
Db    378 TAATGTAAATCAAGACNCAACAATTTCATTTGAATATGGTGTTAGATCCTCT 327


RESULT 9
BJ383574/c
LOCUS       BJ383574                 655 bp    mRNA    linear   EST 08-MAR-2002
DEFINITION  BJ383574 Dictyostelium discoideum cDNA library, CF Dictyostelium
            discoideum cDNA clone ddc53b16 3', mRNA sequence.
ACCESSION   BJ383574
VERSION     BJ383574.1  GI:19292958
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 655)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the culmination
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .655
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddc53b16"
                     /clone_lib="Dictyostelium discoideum cDNA library, CF"
                     /sex="mat A"
                     /dev_stage="Culmination stage"
BASE COUNT     197 a    106 c     95 g    254 t      3 others
ORIGIN

  Query Match            14.9%;  Score 35.8;  DB 13;  Length 655;
  Best Local Similarity  57.1%;  Pred. No. 49;
  Matches   64;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy    125 CTTTGGTGACCTGCCTAAAAATACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGT 184
          | || ||||| |   ||| |||| ||  |    ||| || | ||   || || |||    |
Db    523 CGTTAGTGACTTTGATAATAATAATAATACTNAAAAGAGTTTTGTTGAAAGAGTGATTGG 464

Qy    185 TAATATAAAACAAGAACTAATAAACGCTCTCGCACCCGCAATTATATTTTGT 236
          |||| |||| |||||    || ||    |  | | |    |    ||| ||   | |
Db    463 TAATGTAAATCAAGACACAACAATTTCATTTGAATATGGTGTTAGATCCTCT 412


RESULT 10
AZ526385
LOCUS       AZ526385                 661 bp    DNA     linear   GSS 07-MAY-2001
DEFINITION  254PbH05 Pb MBN #21 Plasmodium berghei genomic 3', DNA sequence.
ACCESSION   AZ526385
VERSION     AZ526385.1  GI:13968158
KEYWORDS    GSS.
SOURCE      Plasmodium berghei.
  ORGANISM  Plasmodium berghei
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 661)
  AUTHORS   Carlton,J.M.-R. and Dame,J.B.
  TITLE     The Plasmodium vivax and P. berghei gene sequence tag projects
  JOURNAL   Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)
COMMENT     Contact: Dame JB
            Dept. of Pathobiology, College of Veterinary Medicine
            University of Florida
            2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
            Tel: 352 392 4700
            Fax: 352 392 9704
            Email: damej@mail.vetmed.ufl.edu
            Seq primer: M13(-20) forward
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .661
                     /organism="Plasmodium berghei"
                     /strain="ANKA clone 15cy1 (clone of the ANKA 8417 clone)"
                     /db_xref="taxon:5821"
                     /clone_lib="Pb MBN #21"
                     /dev_stage="asexual blood forms"
                     /lab_host="Mus musculus"
                     /note="Vector: pBluescript SK(+) vector DNA, phagemid
                     excised from lambda ZAP; Site_1: EcoRV; Site_2: EcoRV;
                     Genomic DNA was prepared from asynchronous blood stage
                     forms of the  cloned ANKA isolate of P. berghei grown in
                     laboratory Swiss white mice.  The DNA was purified from
                     contaminating host DNA by Hoechst Dye  33258-CsCl
                     ultracentrifugation and precipitated. Purified DNA was
                     digested with mung bean nuclease in the presence of 36-38%
                     formamide  at 50 C, as described (Vernick, K.D., Imberski,
                     R.B., and McCutchan, T.F.  1988. Nucleic Acids Research
                     16:6883-6896). The ends of the digestion  fragments were
                     polished using T4 DNA polymerase, and the fragments  size
                     selected in the range 500-2000 bp. These were ligated into
                     the  EcoRV-cleaved and dephosphorylated pBluescript SK(+)
                     vector.  Recombinant plasmids were used to transform E.
                     coli XL10-Gold host cells. "
BASE COUNT     323 a     51 c     98 g    188 t      1 others
ORIGIN

  Query Match            14.8%;  Score 35.6;  DB 17;  Length 661;
  Best Local Similarity  54.6%;  Pred. No. 55;
  Matches   71;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy     85 GAAACCACTGGACGGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAA 144
          ||||  |||       | |||| ||||  |      |||| || |||| |  |    |  || |
Db     40 GAAAGAACTTTTAAAAATTTTGTGATAATATATTTATAGACTTTGAAAAAATCGGAAGA 99

Qy    145 ATACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAACTAA 204
            | || ||||||||    |||| || | |    |        | | | |||| || ||  ||
Db    100 TGGCAACGAACAAAATAGATATTCGCAAAAAAATAATTTATTAAAAAAAAAAAAAAAAAA 159

Qy    205 TAAACGCTCT 214
            | | |  ||
Db    160 AAGAAGACCT 169
```

```
RESULT 11
AL728080/c
LOCUS       AL728080                 524 bp    mRNA    linear   EST 18-APR-2002
DEFINITION  AL728080 Danio rerio embryonic inner ear subtracted cDNA Danio
            rerio cDNA clone BN0AA094ZB11 3', mRNA sequence.
ACCESSION   AL728080
VERSION     AL728080.1  GI:20192684
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 524)
  AUTHORS   Coimbra,R., Weil,D., Brottier,P., Blanchard,S., Levi,M., Hardelin
            ,J.P., Weissenbach,J. and Petit,C.
  TITLE     A subtracted cDNA library from the zebrafish (Danio rerio)
            embryonic inner ear
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .524
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="BN0AA094ZB11"
                     /clone_lib="Danio rerio embryonic inner ear subtracted
                     cDNA"
                     /tissue_type="inner ear"
                     /dev_stage="embryonic"
                     /note="subtracted cDNA library"
BASE COUNT      194 a       68 c      78 g     184 t
ORIGIN

  Query Match             14.8%;  Score 35.4;  DB 9;  Length 524;
  Best Local Similarity   53.2%;  Pred. No. 65;
  Matches   75;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy        100 ATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAACAAAA 159
              |||||| || |   || ||||      ||    || ||| || | ||| |  | | ||| |
Db        318 ATTTTTTTGTTTAGTGCTTTAACCCTTTAACTGCCCTACCAAGAAAAAAAAAAATACATAT 259

Qy        160 TGATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAACTAATAAACGCTCTCGCAC 219
                |      ||   |     ||     | | |   || ||   |||||||| | | | ||
Db        258 GCTATGAAAAAATGAAGATCAATTTTTAAAACATTAAAAAAAAATAAACGTTTTTGAAC 199

Qy        220 CCGCAATTATATTTTGTTCTT 240
                |  ||||||||    || ||
Db        198 ACTAAATTATATTCATTTTTT 178


RESULT 12
BJ356150/c
LOCUS       BJ356150                 772 bp    mRNA    linear   EST 07-MAR-2002
DEFINITION  BJ356150 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda59111 3', mRNA sequence.
ACCESSION   BJ356150
VERSION     BJ356150.1  GI:19255745
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 772)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .772
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda59111"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
BASE COUNT      265 a      125 c     158 g     224 t
ORIGIN

  Query Match             14.8%;  Score 35.4;  DB 13;  Length 772;
  Best Local Similarity   53.2%;  Pred. No. 60;
  Matches   75;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy         98 GGATTTTTATGATGTTTGTTTTAGAGGCTTTGGTGACCTGCCTAAAAATACTACAAACAA 157
              || ||    || || ||      |  ||     | |||||       || ||||  |  ||||||
Db        153 GGCCTTACTTGGTGATTTAGCTCAAGCTATGGGTGATAGTATTAGAAATCTTTTAAACAA 94

Qy        158 AATGATTATTTGCGAGAGCGTGGGCGTTAATATAAAACAAGAACTAATAAACGCTCTCGC 217
                | || ||| | ||      ||     || |   |||   || |||     ||   |  | |||
Db         93 CAAAATAATTAGAGAATTTGTAAACGATGGTATTCAAAAAGGTATACCATATGCTGCATT 34

Qy        218 ACCCGCAATTATATTTTGTTC 238
                |  |||| ||||    || |||
Db         33 ATGCGCATTTATCAATTTTTC 13


RESULT 13
BJ444759/c
LOCUS       BJ444759                 541 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ444759 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv57j09 3', mRNA sequence.
ACCESSION   BJ444759
VERSION     BJ444759.1  GI:19419480
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 541)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .541
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv57j09"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
                     /sex="mat A"
                     /dev_stage="Growth phase"
BASE COUNT      163 a       85 c      81 g     212 t
ORIGIN

  Query Match             14.7%;  Score 35.2;  DB 13;  Length 541;
  Best Local Similarity   57.1%;  Pred. No. 73;
  Matches   64;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        125 CTTTGGTGACCTGCCTAAAAATACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGT 184
```

```
          | || ||||| |   ||| |||| ||  |   ||| || | ||   || || |||   |
Db    502 CGTTAGTGACTTTGATAATAATAATAATACTGAAAAGAGTTTGTTGAAAGAGTGATTGG 443

Qy    185 TAATATAAAACAAGAACTAATAAAACGCTCTCGCACCCGCAATTATATTTTGT 236
          |||| |||| |||||   || ||   |  | | |   |   ||| ||  | |
Db    442 TAATGTAAATCAAGACACAACAATTTCATTTGAATATGGTGTTAGATCCTCT 391


RESULT 14
BJ444864/c
LOCUS       BJ444864                 549 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ444864 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv57n13 3', mRNA sequence.
ACCESSION   BJ444864
VERSION     BJ444864.1  GI:19419585
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..549
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv57n13"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
                     /sex="mat A"
                     /dev_stage="Growth phase"
BASE COUNT      158 a      89 c      81 g     219 t       2 others
ORIGIN

  Query Match           14.7%;  Score 35.2;  DB 13;  Length 549;
  Best Local Similarity   57.1%;  Pred. No. 72;
  Matches   64;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy    125 CTTTGGTGACCTGCCTAAAAATACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGT 184
          | || ||||| |   ||| |||| ||  |   ||| || | ||   || || |||   |
Db    548 CGTTAGTGACTTTGATAATAATAATAATACTGAAAAGAGTTTGTTGAAAGAGTGATTGG 489

Qy    185 TAATATAAAACAAGAACTAATAAACGCTCTCGCACCCGCAATTATATTTTGT 236
          |||| |||| |||||   || ||   |  | | |   |   ||| ||  | |
Db    488 TAATGTAAATCAAGACACAACAATTTCATTTGAATATGGTGTTAGATCCTCT 437


RESULT 15
BJ440866/c
LOCUS       BJ440866                 594 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ440866 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv44p24 3', mRNA sequence.
ACCESSION   BJ440866
VERSION     BJ440866.1  GI:19415588
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 594)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..594
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv44p24"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
                     /sex="mat A"
                     /dev_stage="Growth phase"
BASE COUNT      176 a      97 c      90 g     231 t
ORIGIN

  Query Match           14.7%;  Score 35.2;  DB 13;  Length 594;
  Best Local Similarity   57.1%;  Pred. No. 71;
  Matches   64;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy    125 CTTTGGTGACCTGCCTAAAAATACTACAAACAAAATGATTATTTGCGAGAGCGTGGGCGT 184
          | || ||||| |   ||| |||| ||  |   ||| || | ||   || || |||   |
Db    553 CGTTAGTGACTTTGATAATAATAATAATACTGAAAAGAGTTTGTTGAAAGAGTGATTGG 494

Qy    185 TAATATAAAACAAGAACTAATAAACGCTCTCGCACCCGCAATTATATTTTGT 236
          |||| |||| |||||   || ||   |  | | |   |   ||| ||  | |
Db    493 TAATGTAAATCAAGACACAACAATTTCATTTGAATATGGTGTTAGATCCTCT 442


Search completed: May 12, 2003, 02:35:18
Job time : 2935.71 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 01:02:58 ; Search time 99 Seconds
(without alignments)
5879.355 Million cell updates/sec

Title: US-09-980-277-2
Perfect score: 20
Sequence: 1 gatotgaatccgaccaaccg 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	20	100.0	121	3	SCMRPTAA	M61098 S.mansoni t
2	16.8	84.0	4676	3	DROMYHB	J02788 D.melanogas
3	16.8	84.0	22663	3	DROMHC	M61229 D.melanogas
4	16.8	84.0	68379	3	AC005119	AC005119 Drosophil
5	16.8	84.0	79309	2	AC019767	AC019767 Drosophil
6	16.8	84.0	175413	3	AC093097	AC093097 Drosophil
7	16.8	84.0	258166	3	AE003652	AE003652 Drosophil
8	16.4	82.0	99542	2	AC095629	AC095629 Rattus no
9	16.4	82.0	132927	8	AC079890	AC079890 Oryza sat
c 10	16.4	82.0	145668	9	AC092272	AC092272 Homo sapi
c 11	16.4	82.0	155837	2	AC019350	AC019350 Homo sapi
12	16.4	82.0	165237	9	AC091914	AC091914 Homo sapi
13	16.4	82.0	177246	2	AC114308	AC114308 Homo sapi
c 14	16	80.0	9359	2	AC018248	AC018248 Drosophil
c 15	16	80.0	172069	3	AC012160	AC012160 Drosophil
c 16	16	80.0	177793	3	AC012096	AC012096 Drosophil
c 17	16	80.0	209280	2	AC116528	AC116528 Mus muscu
c 18	16	80.0	298514	3	AE003504	AE003504 Drosophil
c 19	15.8	79.0	355	3	AF449752	AF449752 Sceptodro
c 20	15.8	79.0	754	9	HSA328235	AJ328235 Homo sapi
c 21	15.8	79.0	1655	5	XLVGBEZA1	X56969 X.laevis mR
22	15.8	79.0	3671	6	AR171292	AR171292 Sequence
23	15.8	79.0	3671	6	AR172793	AR172793 Sequence
24	15.8	79.0	10019	1	AE008490	AE008490 Streptoco
25	15.8	79.0	24519	2	AC020211	AC020211 Drosophil
26	15.8	79.0	61009	2	AC084776	AC084776 Homo sapi
c 27	15.8	79.0	68090	2	AC113956	AC113956 Homo sapi
28	15.8	79.0	73397	2	AC101427	AC101427 Mus muscu
29	15.8	79.0	74475	8	T7M24	AP077408 Arabidops
30	15.8	79.0	90632	8	AC021045	AC021045 Arabidops
31	15.8	79.0	94319	8	ATF3L17	AL080283 Arabidops
32	15.8	79.0	103116	2	AC128838	AC128838 Rattus no
c 33	15.8	79.0	103930	9	AC092167	AC092167 Homo sapi
34	15.8	79.0	110000	2	AC119569_2	Continuation (3 of
c 35	15.8	79.0	113721	2	AC130506	AC130506 Rattus no
c 36	15.8	79.0	121542	2	AC122993	AC122993 Rattus no
37	15.8	79.0	122697	9	AC006062	AC006062 Homo sapi
c 38	15.8	79.0	133279	8	OSJN00152	AL662950 Oryza sat
c 39	15.8	79.0	134862	2	AC116433	AC116433 Bos tauru
40	15.8	79.0	135934	2	AC018925	AC018925 Homo sapi
c 41	15.8	79.0	137478	2	AP004704	AP004704 Oryza sat
42	15.8	79.0	142085	2	AC036153	AC036153 Homo sapi
c 43	15.8	79.0	149879	2	AP005381	AP005381 Oryza sat
44	15.8	79.0	150584	2	AC106994	AC106994 Rattus no
45	15.8	79.0	154383	2	AC097312	AC097312 Rattus no

ALIGNMENTS

RESULT 1
SCMRPTAA
LOCUS SCMRPTAA 121 bp DNA linear INV 26-APR-1993
DEFINITION S.mansoni tandem repeat units.
ACCESSION M61098
VERSION M61098.1 GI:161064
KEYWORDS
SOURCE S.mansoni (strain Egyptian) DNA, clone pSml-7.
ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE 1 (bases 1 to 121)
AUTHORS Hamburger,J., Turetski,T., Kapeller,I. and Deresiewicz,R.
TITLE Highly repeated short DNA sequences in the genome of Schistosoma mansoni recognized by a species-specific probe

```
  JOURNAL    Mol. Biochem. Parasitol. 44 (1), 73-80 (1991)
  MEDLINE    91187067
   PUBMED    2011155
FEATURES             Location/Qualifiers
     source          1. .121
                     /organism="Schistosoma mansoni"
                     /db_xref="taxon:6183"
BASE COUNT       36 a      20 c      26 g      39 t
ORIGIN

  Query Match           100.0%;  Score 20;  DB 3;  Length 121;
  Best Local Similarity  100.0%;  Pred. No. 4.4;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GATCTGAATCCGACCAACCG 20
          ||||||||||||||||||||
Db      1 GATCTGAATCCGACCAACCG 20


RESULT 2
DROMYHB
LOCUS       DROMYHB                 4676 bp    DNA     linear   INV 26-APR-1993
DEFINITION  D.melanogaster myosin heavy chain gene, 5' end.
ACCESSION   J02788
VERSION     J02788.1  GI:157939
KEYWORDS    myosin; myosin heavy chain.
SOURCE      D.melanogaster (Canton S) DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 4676)
  AUTHORS   Wassenberg,D.R. II, Kronert,W.A., O'Donnell,P.T. and Bernstein,S.I.
  TITLE     Analysis of the 5' end of the Drosophila muscle myosin heavy chain
            gene. Alternatively spliced transcripts initiate at a single site
            and intron locations are conserved compared to myosin genes of
            other organisms
  JOURNAL   J. Biol. Chem. 262 (22), 10741-10747 (1987)
  MEDLINE   87280141
   PUBMED   3038896
COMMENT     Draft entry and computer-readable copy of sequence in [1] kindly
            provided by S.I.Bernstein, 23-JUN-1987.
FEATURES             Location/Qualifiers
     source          1. .4676
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
     prim_transcript 456. .>4676
                     /note="MHC mRNA"
     intron          472. .2283
                     /note="MHC intron A"
     CDS             join(2381. .2584,2835. .2978,4138. .4518,4592. .4655)
                     /partial
                     /note="myosin heavy-chain"
                     /codon_start=1
                     /protein_id="AAA28706.1"
                     /db_xref="GI:552100"
                     /translation="MPKPVANQEDEDPTPYLPVSLEQRRIDQSKPYDSKKSCWIPDRE
                     EGYLLGEIKATKGDIVSVGLQGGKVRDIKFSKVEKVNPPKFEKIEDMADMTVLNTPCV
                     LHNLRQRYYAKLIYTYSGLFCVAINPYKPVYVTNRCAKMYRGKRENEVPPHIFAISD
                     GAYVDMLTNHVNQSMLITGESGAGKTENTKKVIAYFATVGASKKTDEAAKSKGSMEDQ
                     VVQTNPVLEAFGNAKTVRNDNSSRFGKFIRIHFGPTGKLAGADIET"
     CDS             join(2381. .2584,3144. .3287,4138. .4518,4592. .4655)
                     /partial
                     /note="myosin heavy-chain a"
                     /codon_start=1
                     /protein_id="AAA28707.1"
                     /db_xref="GI:552101"
                     /translation="MPKPVANQEDEDPTPYLPVSLEQRRIDQSKPYDSKKSCWIPDRE
                     EGYLLGEIKATKGDIVSVGLQGGKTRDLKKDLLQQVNPFKYTNAADMSNLTYLNDASV
                     LHNLRQRYYTKLIYTYSGLFCVAINPYKPVYVTNRCAKMYRGKRENEVPPHIFAISD
                     GAYVDMLTNHVNQSMLITGESGAGKTENTKKVIAYFATVGASKKTDEAAKSKGSMEDQ
                     VVQTNPVLEAFGNAKTVRNDNSSRFGKFIRIHFGPTGKLAGADIET"
     exon            <2381. .2584
                     /note="myosin heavy-chain,  (first expressed exon)"
                     /number=2
     exon            <2381. .2584
                     /note="myosin heavy-chain a,  (first expressed exon)"
                     /number=2
     intron          2585. .2834
                     /note="MHC intron B"
     gene            2835. .2978
                     /gene="myosin heavy chain"
     exon            <2835. .2978
                     /gene="myosin heavy chain"
     intron          2979. .3143
                     /note="MHC intron C"
     gene            3144. .3287
                     /gene="myosin heavy chain a"
     exon            3144. .3287
                     /partial
                     /gene="myosin heavy chain a"
     intron          3288. .4137
                     /note="MHC intron D"
     exon            4138. .>4518
                     /note="myosin heavy-chain"
                     /number=4
     exon            4138. .>4518
                     /note="myosin heavy-chain a"
                     /number=4
     intron          4519. .4591
                     /note="MHC intron E"
     exon            4592. .>4655
                     /note="myosin heavy-chain"
                     /number=5
     exon            4592. .>4655
                     /note="myosin heavy-chain a"
                     /number=5
     intron          4656. .>4676
                     /note="MHC intron F"
BASE COUNT     1582 a    984 c    869 g   1241 t
ORIGIN      1 bp upstream of XbaI site.

  Query Match           84.0%;  Score 16.8;  DB 3;  Length 4676;
  Best Local Similarity  90.0%;  Pred. No. 2.7e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GATCTGAATCCGACCAACCG 20
          || |||||||||||| |||||
Db   3106 GAACTGAATCCGACAAACCG 3125


RESULT 3
DROMHC
LOCUS       DROMHC                  22663 bp    DNA     linear   INV 26-APR-1993
DEFINITION  D.melanogaster myosin heavy chain gene, complete cds.
ACCESSION   M61229 M27194
VERSION     M61229.1  GI:157890
KEYWORDS    myosin heavy chain.
SOURCE      D.melanogaster (strain Canton S) DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 22663)
  AUTHORS   George,E.L., Ober,M.B. and Emerson,C.P. Jr.
  TITLE     Functional domains of the Drosophila melanogaster muscle myosin
            heavy-chain gene are encoded by alternatively spliced exons
  JOURNAL   Mol. Cell. Biol. 9 (7), 2957-2974 (1989)
  MEDLINE   89384556
   PUBMED   2506434
FEATURES             Location/Qualifiers
     source          1. .22663
                     /organism="Drosophila melanogaster"
                     /strain="Canton S"
                     /db_xref="taxon:7227"
                     /map="chromosome 36B(2L)"
```

```
exon            456. .471
                /gene="MHC"
                /number=1
exon            2283. .2583
                /gene="MHC"
                /number=2
gene            join(2380. .2583,3146. .3289,4144. .4524,4597. .4660,
                4729. .4827,6968. .7071,8142. .8549,9373. .9543,10538. .11127,
                12078. .12195,14122. .14385,15902. .16111,16333. .17219,
                17391. .17469,18223. .19398,19466. .20371,21578. .21661)
                /gene="MHC"
CDS             join(2380. .2583,3146. .3289,4144. .4524,4597. .4660,
                4729. .4827,5029. .5132,8142. .8549,9373. .9543,10538. .11127,
                13299. .13416,14122. .14385,15902. .16111,16333. .17219,
                17808. .17886,18223. .19398,19466. .20371,21578. .21661)
                /gene="MHC"
                /note="late-pupal stage myosin heavy chain"
                /codon_start=1
                /product="myosin heavy chain"
                /protein_id="AAA28687.1"
                /db_xref="GI:157892"
                /translation="MPKPVANQEDEDPTPYLFVSLEQRRIDQSKPYDSKKSCWIPDEK
                EGYLLGEIKATKGDIVSVGLQGGETRDLKKDLLQQVNPPKYEKAEDMSNLTYLNDASV
                LHNLRQRYYNKLIYTYSGLFCVAINPYKRYPVYTNRCAKMYRGKRRNEVPPHIFAISD
                GAYVDMLTNHVNQSMLITGESGAGKTENTKKVIAYFATVGASKKTDEAAKSKGSLEDQ
                VVQTNPVLEAFGNAKTVRNDNSSRFGKFIRIHFGPTGKLAGADIETYLLEKARVISQQ
                SLERCYHIFYQIMSGSVPGVKDICLLTDNIYDYHIVSQGKVTVASIDDAEEFSLTDQA
                FDILGFTKQEKEDVYRITAAVMHMGGMKFKQRGREEQAEQDGEEEGGRVSKLFGCDTA
                ELYKNLLKPRIKVGNEFVTQGRNVQQVTNSIGALCKGVFDRLFKWLVKKCNETLDTQQ
                KRQHFIGVLDIAGFEIFEYNGFEQLCINFTNEKLQQFFNHHMFVLEQEEYKREGIDWA
                FIDFGMDLLACIDLIEKPMGILSILEEESMFPKATDQTFSEKLTNTHLGKSAPFQKPK
                PPKPGQQAAHFAIAHYAGCVSYNITGWLEKNKDPLNDTVVDQFKKSQNKLLIEIFADH
                AGQSGGGEQAKGGRGKKGGGFATVSSAYKEQLNSLMTTLRSTQPHFVRCIIPNEMKQP
                GVVDAHLVMHQLTCNGVLEGIRICRKGFPNRMMYPDFKMRYKIMCPKLLQGVEKDKKA
                TEIIIKFIDLPEDQYRLGNTKVFFRAGVLGQMEEFRDERLGKIMSWMQAWARGYLSRK
                GFKKLQEQRVALKVVQRNLRKYLQLRTWPWYKLWQKVKPLLNVSRIEDEIARLEEKAK
                KAEELHAAEVKVRKELEALNAKLLAEKTALLDSLSGEKGALQDYQERNAKLTAQKNDL
                ENQLRDIQERLTQEEDARNQLFQQKKKADQEISGLKKDIEDLELNVQKAEQDKATKDH
                QIRNLNDEIAHQDELINKLNKEKKMQGETNQKTGEELQAAEDKINHLNKVKAKLEQTL
                DELEDSLEREKKVRGDVEKSKRKVEGDLKLTQEAVADLERNKKELEQTIQRKDKELSS
                ITAKLEDEQVVVLKHQRQIKELQARIEELEEEVEAERQARAKAEKQRADLARELEELG
                ERLEEAGGATSAQIELNKKREAELSKLRRDLEEANIQHESTLANLRKKHNDAVAEMAE
                QVDQLNKLKAKAEKEKNEYYGQLNDLRAGVDHITNEKAAQEKIAKQLQHTLNEVQSKL
                DETNRTLNDFDASKKKLSIENSDLLRQLEEAESQVSQLSKIKISLTTQLEDTKRLADE
                ESRERATLLGKFRNLEHDLDNLREQVEEEAEGKADLQRQLSKANAEAQVWRSKYESDG
                VARSEELEEAKRKLQARLAEAEETIESLNQKCIGLEKTKQRLSTEVEDLQLEVDRANA
                IANAAEKKQKAFDKIIGEWKLKVDDLAAELDASQKECRNYSTELFRLKGAYEEGQEQL
                EAVRRENKNLADEVKDLLDQIGEGGRNIHEIEKARKRLEAEKDELQAALEEAEAALEQ
                EENKVLRAQLELSQVRQEIDRRIQEKEEEFENTRKNHQRALDSMQASLEAEAKGKAEA
                LRMKKKLEADINELEIALDHANKANAEAQKNIKRYQQQLKDIQTALEEEQRARDDARE
                QLGISERRANALQNELEESRTLLEQADRGRRQAEQELADAHEQLNEVSAQNASISAAK
                RKLESELQTLHSDLDELLNEAKNSEEKAKKAMVDAARLADELRAEQDHAQTQEKLRKA
                LEQQIKELQVRLDEAEANALKGGKKAIQKLEQRVRELENELDGEQRRHADAQKNLRKS
                ERRVKELSFQSEEDRKNHERMQDLVDKLQQKIKTYKRQIEEAEEIAALNLAKFRKAQQ
                ELEEAEERADLAEQAISKFRAKGRAGSVGRGASPAPRATSVRPQFDGLAFPPRFDLAP
                ENEF"
CDS             join(2380. .2583,3146. .3289,4144. .4524,4597. .4660,
                4729. .4827,6968. .7071,8142. .8549,9373. .9543,10538. .11127,
                12078. .12195,14122. .14385,15902. .16111,16333. .17219,
                17391. .17469,18223. .19398,19466. .20371,21578. .21661)
                /gene="MHC"
                /note="late-pupal stage myosin heavy chain"
                /codon_start=1
                /product="myosin heavy chain"
                /protein_id="AAA28686.1"
                /db_xref="GI:157891"
                /translation="MPKPVANQEDEDPTPYLFVSLEQRRIDQSKPYDSKKSCWIPDEK
                EGYLLGEIKATKGDIVSVGLQGGETRDLKKDLLQQVNPPKYEKAEDMSNLTYLNDASV
                LHNLRQRYYNKLIYTYSGLFCVAINPYKRYPVYTNRCAKMYRGKRRNEVPPHIFAISD
                GAYVDMLTNHVNQSMLITGESGAGKTENTKKVIAYFATVGASKKTDEAAKSKGSLEDQ
                VVQTNPVLEAFGNAKTVRNDNSSRFGKFIRIHFGPTGKLAGADIETYLLEKARVISQQ
                SLERCYHIFYQIMSGSVPGVKEMCFLSDNIYDYYNVSQGKVTVPNMDDGEEFQLADQA
                FDILGFTKQEKEDVYRITAAVMHMGGMKFKQRGREEQAEQDGEEEGGRVSKLFGCDTA
                ELYKNLLKPRIKVGNEFVTQGRNVQQVTNSIGALCKGVFDRLFKWLVKKCNETLDTQQ
                KRQHFIGVLDIAGFEIFEYNGFEQLCINFTNEKLQQFFNHHMFVLEQEEYKREGIDWA
                FIDFGMDLLACIDLIEKPMGILSILEEESMFPKATDQTFSEKLTNTHLGKSAPFQKPK
                PPKPGQQAAHFAIAHYAGCVSYNITGWLEKNKDPLNDTVVDQFKKSQNKLLIEIFADH
                AGQSGGGEQAKGGRGKKGGGFATVSSAYKEQLNSLMTTLRSTQPHFVRCIIPNEMKQP
                GVVDAHLVMHQLTCNGVLEGIRICRKGFPNRMMYPDFKMRYQILNPAGIVGVDDPKKC
                GSIILESTALDPDMYRIGHTKVFFRAGVLGQMEEFRDERLGKIMSWMQAWARGYLSRK
                GFKKLQEQRVALKVVQRNLRKYLQLRTWPWYKLWQKVKPLLNVSRIEDEIARLEEKAK
                KAEELHAAEVKVRKELEALNAKLLAEKTALLDSLSGEKGALQDYQERNAKLTAQKNDL
                ENQLRDIQERLTQEEDARNQLFQQKKKADQEISGLKKDIEDLELNVQKAEQDKATKDH
                QIRNLNDEIAHQDELINKLNKEKKMQGETNQKTGEELQAAEDKINHLNKVKAKLEQTL
                DELEDSLEREKKVRGDVEKSKRKVEGDLKLTQEAVADLERNKKELEQTIQRKDKELSS
                ITAKLEDEQVVVLKHQRQIKELQARIEELEEEVEAERQARAKAEKQRADLARELEELG
                ERLEEAGGATSAQIELNKKREAELSKLRRDLEEANIQHESTLANLRKKHNDAVAEMAE
                QVDQLNKLKAKAEHDRQTCHNELNQTRTACDQLGRDKAAQEKIAKQLQHTLNEVQSKL
                DETNRTLNDFDASKKKLSIENSDLLRQLEEAESQVSQLSKIKISLTTQLEDTKRLADE
                ESRERATLLGKFRNLEHDLDNLREQVEEEAEGKADLQRQLSKANAEAQVWRSKYESDG
                VARSEELEEAKRKLQARLAEAEETIESLNQKCIGLEKTKQRLSTEVEDLQLEVDRANA
                IANAAEKKQKAFDKIIGEWKLKVDDLAAELDASQKECRNYSTELFRLKGAYEEGQEQL
                EAVRRENKNLADEVKDLLDQIGEGGRNIHEIEKARKRLEAEKDELQAALEEAEAALEQ
                EENKVLRAQLELSQVRQEIDRRIQEKEEEFENTRKNHQRALDSMQASLEAEAKGKAEA
                LRMKKKLEADINELEIALDHANKANAEAQKNIKRYQQQLKDIQTALEEEQRARDDARE
                QLGISERRANALQNELEESRTLLEQADRGRRQAEQELADAHEQLNEVSAQNASISAAK
                RKLESELQTLHSDLDELLNEAKNSEEKAKKAMVDAARLADELRAEQDHAQTQEKLRKA
                LEQQIKELQVRLDEAEANALKGGKKAIQKLEQRVRELENELDGEQRRHADAQKNLRKS
                ERRVKELSFQSEEDRKNHERMQDLVDKLQQKIKTYKRQIEEAEEIAALNLAKFRKAQQ
                ELEEAEERADLAEQAISKFRAKGRAGSVGRGASPAPRATSVRPQFDGLAFPPRFDLAP
                ENEF"
exon            2774. .2977
                /gene="MHC"
exon            3146. .3289
                /gene="MHC"
exon            4144. .4524
                /gene="MHC"
                /number=4
exon            4597. .4660
                /gene="MHC"
                /number=5
exon            4729. .4827
                /gene="MHC"
                /number=6
exon            5029. .5132
                /gene="MHC"
exon            5626. .5729
                /gene="MHC"
exon            6121. .6224
                /gene="MHC"
exon            6968. .7071
                /gene="MHC"
exon            8142. .8549
                /gene="MHC"
                /number=8
exon            8902. .9072
                /gene="MHC"
exon            9373. .9543
                /gene="MHC"
exon            9951. .10121
                /gene="MHC"
exon            10538. .11127
                /gene="MHC"
                /number=10
exon            11319. .11436
                /gene="MHC"
exon            11693. .11810
                /gene="MHC"
exon            12078. .12195
                /gene="MHC"
exon            12663. .12781
                /gene="MHC"
exon            13299. .13416
                /gene="MHC"
exon            14122. .14385
                /gene="MHC"
```

```
                     /number=12
     exon            15902. .16111
                     /gene="MEC"
                     /number=13
     exon            16333. .17219
                     /gene="MEC"
                     /number=14
     exon            17391. .17469
                     /gene="MEC"
     exon            17808. .17886
                     /gene="MEC"
     exon            18223. .19398
                     /gene="MEC"
                     /number=16
     exon            19466. .20371
                     /gene="MEC"
                     /number=17
     exon            20636. .21137
                     /gene="MEC"
                     /number=18
     exon            21578. .22322
                     /gene="MEC"
                     /number=19
     exon            21578. .21771
                     /gene="MEC"
                     /number=19
     polyA_signal    21753. .21758
                     /gene="MEC"
     polyA_signal    21831. .21836
                     /gene="MEC"
     polyA_signal    22304. .22309

  Query Match             84.0%;  Score 16.8;  DB 3;  Length 22663;
  Best Local Similarity   90.0%;  Pred. No. 2.6e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACCG 20
           || |||||||||| ||||
Db    3108 GAACTGAATCCGACAAACCG 3127


RESULT 4
AC005119
LOCUS       AC005119               68379 bp    DNA     linear   INV 09-JUL-1999
DEFINITION  Drosophila melanogaster, chromosome 2L, region 36B1-36B2, P1 clone
            DS06436, complete sequence.
ACCESSION   AC005119
VERSION     AC005119.9  GI:5430743
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 68379)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Ling,H.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Richards,S., Sequeira,A., Sethi,N.,
            Svirskas,R.R., Wan,K.H., Webster,D., Woolley,P., Yang,S., Yee,M.,
            Zieran,L.L. and Rubin,G.M.
  TITLE     Sequencing of Drosophila chromosome 2L, region 36B1-36B2
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 68379)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
            Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mak,J., Marzo,P., Mok,M.S., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
            Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
            Zieran,L.L. and Kimmel,B.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-1998) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Jul 9, 1999 this sequence version replaced gi:4680763.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
            P1 library location: 68-4.
FEATURES             Location/Qualifiers
     source          1..68379
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
                     /map="36B1-36B2"
                     /clone="P1 DS06436 (D314)"
                     /clone_lib="P1 library, partial Sau3A in pNS582tet14d10"
                     /note="DS06436 (D314) extends from a minimal overlap with
                     its distal neighbor DS00592 (D315) at bp 1 to P1 end at bp
                     68378."
BASE COUNT     19786 a  15275 c  14103 g  19215 t
ORIGIN

  Query Match             84.0%;  Score 16.8;  DB 3;  Length 68379;
  Best Local Similarity   90.0%;  Pred. No. 2.6e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACCG 20
           || |||||||||| ||||
Db   44948 GAACTGAATCCGACAAACCG 44967


RESULT 5
AC019767
LOCUS       AC019767               79309 bp    DNA     linear   HTG 03-JAN-2000
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC019767
VERSION     AC019767.1  GI:6665130
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 79309)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CRM:10210880 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..79309
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     23312 a  16461 c  15838 g  23698 t
ORIGIN

  Query Match             84.0%;  Score 16.8;  DB 2;  Length 79309;
  Best Local Similarity   90.0%;  Pred. No. 2.6e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;
```

```
Qy      1 GATCTGAATCCGACCAACCG 20
          || ||||||||||| |||||
Db  74946 GAACTGAATCCGACAAACCG 74965


RESULT 6
AC093097
LOCUS       AC093097               175413 bp    DNA     linear   INV 09-AUG-2001
DEFINITION  Drosophila melanogaster, chromosome 2L, region 36A-36B, BAC clone
            BACR26003, complete sequence.
ACCESSION   AC093097
VERSION     AC093097.1  GI:15145638
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 175413)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 2L, region 36A-36B
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175413)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2001) Berkeley Drosophila Genome Project, MS
            64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES    Location/Qualifiers
     source 1..175413
            /organism="Drosophila melanogaster"
            /strain="y; cn bw sp"
            /db_xref="taxon:7227"
            /chromosome="2L"
            /map="36A-36B"
            /clone="BACR26003 (D1298)"
            /clone_lib="RPCI-98 (Roswell Park Cancer Institute
            Drosophila melanogaster BAC library, partial EcoRI in
            pBAC93.6)"
BASE COUNT     52084 a  36665 c  35604 g  51060 t
ORIGIN

  Query Match             84.0%;  Score 16.8;  DB 3;  Length 175413;
  Best Local Similarity   90.0%;  Pred. No. 2.6e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GATCTGAATCCGACCAACCG 20
          || ||||||||||| |||||
Db 142511 GAACTGAATCCGACAAACCG 142530


RESULT 7
AE003652
LOCUS       AE003652               258166 bp    DNA     linear   INV 04-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386055 section 45
            of 63, complete sequence.
ACCESSION   AE003652 AE002690
VERSION     AE003652.1  GI:7298316
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 258166)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Sidén-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
  PUBMED    10731132
REFERENCE   2  (bases 1 to 258166)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
```

```
                Rockville, MD, USA
FEATURES        Location/Qualifiers
     source     1. .258166
                /organism="Drosophila melanogaster"
                /db_xref="taxon:7227"
                /chromosome="2L"
     gene       <6056. .>9319
                /gene="CG13261"
                /evidence=not_experimental
                /map="36A5-36A5"
                /db_xref="FLYBASE:FBan0013261"
                /db_xref="FLYBASE:FBgn0032591"
     mRNA       join(<6056. .6965,7234. .7285,9202. .>9319)
                /gene="CG13261"
                /product="CT32545"
                /evidence=not_experimental
                /db_xref="FLYBASE:FBan0013261"
                /db_xref="FLYBASE:FBgn0032591"
     CDS        join(6056. .6965,7234. .7285,9202. .9319)
                /gene="CG13261"
                /note="CG13261 gene product"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAF53545.1"
                /db_xref="GI:7298317"
                /db_xref="FLYBASE:FBan0013261"
                /db_xref="FLYBASE:FBgn0032591"
                /translation="MYRLMVAEQSSSSNNNGNISNSSISNFNGTIITSNNNAGSCSS
                TTSSTTSLNSSNPASPACTMVLPAPQQHLLVQQQQQQQQQQQQQPQQPPQLQQQH
                GPQQQQQQQQQHSPSDGSSNCSLPASPPLLQQTATPPQGAQIVPPVCALHHPQQQL
                ALMAAMQHHHLPPPHALHHAPLPPPPPMLPPPHENGHGGGSGVAVGGGIISPNGG
                GSASSYSESLSSGSGSSGGQVWGLSNGYWIAPLFPQQLTIQTPGEPYPPEYIAPHP
                HEGICPQPHPHHPHHPQPMCAMSTEYDISLSLRHDDAVVELLSCDFPDSFAFTLSSN
                RLWPSLIVVNRINLQRKWALSACWA"
     gene       <36863. .>45904
                /gene="CG4668"
                /map="36A6-36A5"
                /db_xref="FLYBASE:FBan0004668"
                /db_xref="FLYBASE:FBgn0032592"
     mRNA       join(36863. .37038,41625. .42078,42150. .>45904)
                /gene="CG4668"
                /product="CT15043"
                /db_xref="FLYBASE:FBan0004668"
                /db_xref="FLYBASE:FBgn0032592"
     CDS        join(36973. .37038,41625. .42078,42150. .45904)
                /gene="CG4668"
                /note="CG4668 gene product [alt 2]"
                /codon_start=1
                /protein_id="AAF53546.2"
                /db_xref="GI:10728824"
                /db_xref="FLYBASE:FBan0004668"
                /db_xref="FLYBASE:FBgn0032592"
                /translation="MLSPQQYQQLHPQQGHLHPAPPISANGTSHFYTPVAGYPGPGPA
                GNGPPHYHLPPPPPQQQQQQPPQQPQQAPQGPPPPQQQQQQQQVAPPAVQGAPST
                GSYGAAQGAPATAPTHTHAHSHSHSHAHSHTHPQPHAHSPHSPSPNYKDERSQRQR
                NKLLKKLEKQESNPPHSPSPRANLNGHNNNIPPVPLPGHLNNHHAVNHQGRR
                LPQQHGQGAHQVACQQQQSQARNGNPQHPQRAGSSVGASSVGTSEDGDNSSLAGDDE
                EEYHRNSIIEQISAIEKPSYIDVTSRAAKIIWESPAIANTVTVDMRQLRYQVLLCDS
                GKQCKYKSLYQGFAVDCIVQDLPQGDTLVRLQVHTQKLTGTVSDPTEFSTPPCFEPQ
                PPPPKLVSRTKNSINLRWAAPAANGASIQHYLLSEYDEGRMPGQPQRVFELAIKAKHY
                VIGKLQPTTVYSFRIAAVNEAGQSPTSPVASYSTSGNPPPVPKPPQLLASSSSSLKLG
                WEKRAQGDPLLQLQDAETGHGFLNTIKGTELQTECLQLRRASSTQPHLRSENEAGFS
                FWSFEVSTKLAEPGPRGPGKPHAKGKIHGTQFKAAWDAFSDSGSAELLCYHLELSAAG
                PAFERIYSGASTEAMCERLQPGTTYALKACCEGPAQGSPYSDIGHVTTEAVPPSAPPP
                PHCSDPPSFYAALLKLRPEYNGSAPILSFEAQMRELQVQPPQLVTYGKQTYFVAQD
                LTPGGMTEQVRAINEIGAGSWSQMNFTAAAAPQVPEELWVLWSAPHLGVSNQPP
                LQENGAFVTSTTLKSASQERKEDDSEEEVKEPPSSEPHNCYQPQTCAELRNLSPTTR
                YHFRIQASNSAGTGNSSDVISVCTPAAVPGAPQMQGYERTAQEVTLNWTQPAAHSPI
                CEYNIYGKETAPPACTRYTVSGLMPGTGYKFRVQAVNAIGAGAFSAYAKLTTQPA
                PPAPPRLECSGAGHNTIKLKWGENGNGKVVNSPSGNGDFTKFVEMTVRAEQPQA
                VTSGTNCMCKVHKLQERSSYTFRIYAHTDRAGDGTSEEFVETSATLPANIKPPRVV
                QRGSVCLMELPQQLGMQLYLEWQHSKNSFNDRVEYELQYAVLGAAELEGESLSPKGRS
                SSSGGSSNGAPILNPHDYFQLIKRGFETYFPTIDNLAAGTCYQFRVCPVRIAAGEGLLY
```

```
                GQPSSPLRVQVPSELDPSSATCHHHLHGSTATPAMASSQRSSRKLSAGNGSTNGLRMR
                SYSASAISGASGVGADPVAIGRLQQELSGFNACADPQHHHHHHHHHHPQCQGGVGGA
                GGISGATDSHHQHQHHMHSHSHMHHHAHHPHTGMGGSSSNSSSTAAAISSSLAQAGGL
                RRIVCKVTSLYSNRRRLSDQQKAVCIVVSFLVGFFLVAMLVNMLRG"
     mRNA       join(41271. .41555,41625. .42078,42150. .>45904)
                /gene="CG4668"
                /product="CT15049"
                /db_xref="FLYBASE:FBan0004668"
                /db_xref="FLYBASE:FBgn0032592"
     CDS        join(41523. .41555,41625. .42078,42150. .45904)
                /gene="CG4668"
                /note="CG4668 gene product [alt 1]"
                /codon_start=1
                /protein_id="AAF53547.2"
                /db_xref="GI:10728825"
                /db_xref="FLYBASE:FBan0004668"
                /db_xref="FLYBASE:FBgn0032592"
                /translation="MDGSKKTVLLIKISANGTSHFTPVAGYPGPGPAGNGPPHYHLPP
                PPPPQQQQPPQQPQQAPQGPPPPQQQQQQQQVAPPAVQGAPSTGSYGAAQGAPA
                TATATTHTHAHSHSHSHAHSHTHPQPHAHSPHSPSPNYRDERSQRQHNKLLKKLEKQR
                ESNPPHSPSPRANELNGHNNNNIPPGVPLPGHLNNHHAVNHQGRKLPQQHGQGAHQ
                VAQQQQQSQARNGNPQHPQRAGSSVGGASSVGTSEDGDNSSLAGDDEEEYHRNSIIE
                QISAIEKPEVIDVTSRAAKIIWESPAIANTVTVDMRQLRYQVLLCDSGKQCKYKSLYQ
                GEAFECIVQDLPQGQDTLVRLQVHTQKLTGTVSDPTEFTPPCFEPDQPPPKLVSRTK
                NSINLRWAAPAANGASIQHYLLSEYDEGRMPGQPQRVFELAIKAKHYVIGKLQPTTVY
                SFRIAAVNEAGQSPYSPVASYSTSGNPPPVPKPPQLLASSSSSLKLGWERKAQDGDPL
                LQLQDAETGHGFLNTYKGTELQTECLQLRRASSYQPHLRSENEAGFSPWSPEVSYSRTL
                AEPGPRGPGKPHAKGKIHGTQFKAWDAPSDSGGAELCYHLELSAAGPAFERIYSGAS
                TEAMCERLQPGTTYALRACCEGPAQGSPYSDIGPVTTEAVPPSAPPPPHCSDPPSFYA
                ALLKLRPPEYNGSAPILSFEAQMRRLEQVQPPQLVYRGKQTYFVAQDLTPGGMYETQV
                RAINKIGAGSWSQMWFTAAAAPGVPEELWVLVKSATHLGVSWQPPLQENGAPVTST
                TLKSASQERKEDDSEEEVKEPPSSEPHNCYQPQTCAELRNLSPTTRYHFRIQASNSA
                GTGNSSDVISVCTPAAVPGAPQMQGYEPTAQEVTLNWTQPAAHGSPICSYNIRYGERT
                IATPDACTRYTVSGLMPGTGYKFRVQAVNAIGAGAFSAYAKLTTQPAPPAPPRLECSG
                AGHNTIKLKWGENGNGKVVNSPSGNGDFTKFVEMTVRAREQPQAVTSGTNCMCKV
                HKLQERSSYTFRIYAHTDRAGDGDYSEEFVETSATLPANIKPPRVVVQSGSVCLMELP
                QQLGMQLTLEWQHSKNSFNDRVEYELQYAVLGAAELEGESLSPKGRSSSSGGSSNGAP
                ILNPHDYFQLIKRGPETKFTIDNLAAGTCYQFKVCPVRIAAGGEGLLYGQPSSPLRVQV
                PSELDPSSATCHHHLHGSTATPAMASSQRSSRKLSAGNGSTNGLRMRSYSASAISGAS
                GVGADPVAIGRLQQELSGFNACADPQHHHHHHHHHQPCGGVGGAGGISGATDSHH
                QHQHHMHSHSHMHHAHHPHTGMGGSSSNSSSTAAAISSSLAQAGGLRRIVCKVTSLY
                SNRRRLSDQQKAVCIVVSFLVGFFLVAMLVNMLRG"
     gene       complement(<49011. .>53845)
                /gene="trpgamma"
                /note="CG5996"
                /db_xref="FLYBASE:FBan0005996"
     mRNA       complement(join(<49011. .49235,49287. .49405,49467. .49742,
                50044. .50409,50481. .50706,50775. .50969,51041. .51236,
                51293. .51409,51578. .51846,51903. .52198,52268. .52502,
                52561. .52743,52806. .53138,53195. .53563,53624. .>53845))
                /gene="trpgamma"
                /product="CT13821"
                /db_xref="FLYBASE:FBan0005996"
                /db_xref="FLYBASE:FBgn0032593"
     CDS        complement(join(49011. .49235,49287. .49405,49467. .49742,
                50044. .50409,50481. .50706,50775. .50969,51041. .51236,
                51293. .51409,51578. .51846,51903. .52198,52268. .52502,
                52561. .52743,52806. .53138,53195. .53563,53624. .53845))
                /gene="trpgamma"
                /note="trpgamma gene product"
                /codon_start=1
                /protein_id="AAF53548.1"
                /db_xref="GI:7298320"
                /db_xref="FLYBASE:FBan0005996"
                /db_xref="FLYBASE:FBgn0032593"
                /translation="MLLIPGICFLKCNLDDGKKLHFVSPARRKSVLAPILNKIKNGAD
                AADSAEKKLGSSSSGSCHIAVQIEPARRVKRHSIHGMMEEENTIRPHQBIRQLTLEEK
                KFLLAVERGDMAGTRRMLQKAQDTEYINVCVDPLGKTALLMAIDNENLEMVELLINY
```

```
Query Match          84.0%;  Score 16.8;  DB 3;  Length 258166;
Best Local Similarity  90.0%;  Pred. No. 2.6e+02;
Matches  18;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;
```

```
Qy          1 GAPCTGAATCCGACCAACCG 20
              || ||||||||||| |||||
Db     153544 GAACTGAATCCGACAAACCG 153563

RESULT 8
AC095629
LOCUS       AC095629              99542 bp    DNA     linear   HTG 11-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-8F10, *** SEQUENCING IN PROGRESS ***,
            50 unordered pieces.
ACCESSION   AC095629
VERSION     AC095629.3  GI:21722685
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 99542)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Cartier,M., Cavazoz,S.R., Chacko,J., Chaver,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dedersich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,F., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Haviak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joodah,S.,
            Karlsson,E., Kelly,S., Khan,U., Kimball,L., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,M., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oraguuye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 99542)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 99542)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 10, 2002 this sequence version replaced gi:17942183.
            -------------- Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
               Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
               Center project name: GCWA
               Center clone name: CH230-8F10
            -------------- Summary Statistics
               Sequencing vector: Plasmid;
               Chemistry: Dye-terminator Big Dye: 100% of reads
               Assembly program: Phrap; version 0.990329
               Consensus quality: 43940 bases at least Q40
               Consensus quality: 47263 bases at least Q30
               Consensus quality: 50225 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 50 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1745: contig of 1745 bp in length
            *    1746     1845: gap of unknown length
            *    1846     3299: contig of 1454 bp in length
            *    3300     3399: gap of unknown length
            *    3400     4667: contig of 1268 bp in length
            *    4668     4767: gap of unknown length
            *    4768     6262: contig of 1495 bp in length
            *    6263     6362: gap of unknown length
            *    6363     7679: contig of 1317 bp in length
            *    7680     7779: gap of unknown length
            *    7780     9022: contig of 1243 bp in length
            *    9023     9122: gap of unknown length
            *    9123    10340: contig of 1218 bp in length
            *   10341    10440: gap of unknown length
            *   10441    11504: contig of 1064 bp in length
            *   11505    11604: gap of unknown length
            *   11605    13075: contig of 1471 bp in length
            *   13076    13175: gap of unknown length
            *   13176    14566: contig of 1391 bp in length
            *   14567    14666: gap of unknown length
            *   14667    16154: contig of 1488 bp in length
            *   16155    16254: gap of unknown length
            *   16255    17749: contig of 1495 bp in length
            *   17750    17849: gap of unknown length
            *   17850    19505: contig of 1656 bp in length
            *   19506    19605: gap of unknown length
            *   19606    21176: contig of 1571 bp in length
            *   21177    21276: gap of unknown length
            *   21277    22846: contig of 1570 bp in length
            *   22847    22946: gap of unknown length
            *   22947    24186: contig of 1240 bp in length
            *   24187    24286: gap of unknown length
            *   24287    25955: contig of 1669 bp in length
            *   25956    26055: gap of unknown length
            *   26056    27445: contig of 1390 bp in length
            *   27446    27545: gap of unknown length
            *   27546    28570: contig of 1025 bp in length
            *   28571    28670: gap of unknown length
            *   28671    29844: contig of 1174 bp in length
            *   29845    29944: gap of unknown length
            *   29945    31571: contig of 1627 bp in length
            *   31572    31671: gap of unknown length
            *   31672    33601: contig of 1930 bp in length
            *   33602    33701: gap of unknown length
            *   33702    34808: contig of 1107 bp in length
```

```
            *    34809    34908: gap of unknown length
            *    34909    36643: contig of 1735 bp in length
            *    36644    36743: gap of unknown length
            *    36744    38374: contig of 1631 bp in length
            *    38375    38474: gap of unknown length
            *    38475    40687: contig of 2213 bp in length
            *    40688    40787: gap of unknown length
            *    40788    42053: contig of 1266 bp in length
            *    42054    42153: gap of unknown length
            *    42154    43382: contig of 1229 bp in length
            *    43383    43482: gap of unknown length
            *    43483    45071: contig of 1589 bp in length
            *    45072    45171: gap of unknown length
            *    45172    47004: contig of 1833 bp in length
            *    47005    47104: gap of unknown length
            *    47105    48451: contig of 1347 bp in length
            *    48452    48551: gap of unknown length
            *    48552    49892: contig of 1341 bp in length
            *    49893    49992: gap of unknown length
            *    49993    51860: contig of 1868 bp in length
            *    51861    51960: gap of unknown length
            *    51961    53979: contig of 2019 bp in length
            *    53980    54079: gap of unknown length
            *    54080    55641: contig of 1562 bp in length
            *    55642    55741: gap of unknown length
            *    55742    58182: contig of 2441 bp in length
            *    58183    58282: gap of unknown length
            *    58283    61300: contig of 3018 bp in length
            *    61301    61400: gap of unknown length
            *    61401    62816: contig of 1416 bp in length
            *    62817    62916: gap of unknown length
            *    62917    64943: contig of 2027 bp in length
            *    64944    65043: gap of unknown length
            *    65044    67311: contig of 2268 bp in length
            *    67312    67411: gap of unknown length
            *    67412    69802: contig of 2391 bp in length
            *    69803    69902: gap of unknown length
            *    69903    73494: contig of 3592 bp in length
            *    73495    73594: gap of unknown length
            *    73595    76536: contig of 2942 bp in length
            *    76537    76636: gap of unknown length
            *    76637    79951: contig of 3315 bp in length
            *    79952    80051: gap of unknown length
            *    80052    82994: contig of 2943 bp in length
            *    82995    83094: gap of unknown length
            *    83095    85941: contig of 2847 bp in length
            *    85942    86041: gap of unknown length
            *    86042    88827: contig of 2786 bp in length
            *    88828    88927: gap of unknown length
            *    88928    92287: contig of 3360 bp in length
            *    92288    92387: gap of unknown length
            *    92388    94886: contig of 2499 bp in length
            *    94887    94986: gap of unknown length
            *    94987    99542: contig of 4556 bp in length.
FEATURES             Location/Qualifiers
     source          1. .99542
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-8F10"
BASE COUNT     26342 a  21077 c  19949 g  27095 t    5079 others
ORIGIN

  Query Match          82.0%;  Score 16.4;  DB 2;  Length 99542;
  Best Local Similarity  94.4%;  Pred. No. 4.3e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 ATCTGAATCCGACCAACC 19
           |||||||||||| ||||
Db   83403 ATCTGAATCCGAGCAACC 83420


RESULT 9
AC079890
```

```
LOCUS       AC079890               132927 bp    DNA     linear   PLN 29-AUG-2001
DEFINITION  Oryza sativa chromosome 10 BAC OSJNBb0089A17 genomic sequence,
            complete sequence.
ACCESSION   AC079890
VERSION     AC079890.13  GI:13491223
KEYWORDS    HTG.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 132927)
  AUTHORS   Buell,C.R., Yuan,Q., Ouyang,S., Moffat,K.S., Hill,J.N., Burr,P.C.,
            Hsiao,J., Eismann,V., Pai,G., Bowman,C.L., Fujii,C.Y.,
            Vasakos,S.E., Bowman,C.L., Craven,B., Utterback,T.R., Khalak,H.,
            Feldblyum,T.V., Quackenbush,J., White,O., Salzberg,S.L. and
            Fraser,C.M.
  TITLE     Oryza sativa chromosome 10 BAC OSJNBb0089A17 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 132927)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 132927)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   4  (bases 1 to 132927)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
REFERENCE   5  (bases 1 to 132927)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
REFERENCE   6  (bases 1 to 132927)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
COMMENT     On Mar 31, 2001 this sequence version replaced gi:13310897.
            Address all correspondence to:rice@tigr.org

            BAC clone OSJNBb0089A17 is from Oryza sativa chromosome 10
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of several methods: Gene
            prediction programs including genscan and Genscan+ (Chris Burge,
            http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky,
            http://genemark.biology.gatech.edu/GeneMark/), Fgenesh
            (http://www.softberry.com/), and GeneSplicer (Mihaela Pertea and
            Steven Salzberg, contact mpertea@tigr.org), searches of the
            complete sequence against a peptide database and the plant EST
            database at TIGR (http://www.tigr.org/tdb/tgi.shtml). Annotated
            genes are named to indicate the level of evidence for their
            annotation. Genes with similarity to other proteins are named after
            the database hits. Genes without significant peptide similarity but
            with EST similarity are named as  unknown  proteins. Genes without
            protein or EST similarity, that are predicted by more than two gene
            prediction programs over most of their length are annotated as
            hypothetical  proteins. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
            Simple repeats are identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES             Location/Qualifiers
     source          1. .132927
                     /organism="Oryza sativa"
                     /cultivar="Nipponbare"
                     /sub_species="japonica"
```

```
                     /db_xref="taxon:4530"
                     /chromosome="10"
                     /map="near C797"
                     /clone="OSJNBb0089A17"
     repeat_region   complement(3708. .3735)
                     /rpt_family="AT_rich"
     repeat_region   complement(8219. .8252)
                     /rpt_family="(GAAAA)n"
     repeat_region   complement(10188. .10210)
                     /rpt_family="AT_rich"
     repeat_region   11133. .11176
                     /rpt_family="(GAA)n"
     repeat_region   11215. .11237
                     /rpt_family="(GAAAA)n"
     repeat_region   complement(11764. .11796)
                     /rpt_family="AT_rich"
     repeat_region   11867. .11918
                     /rpt_family="(GAAAA)n"
     gene            complement(11962. .12830)
                     /gene="OSJNBb0089A17.22"
                     /note="similar to transcriptional activator CBF1
                     GB:AAC49662 GI:1899058 (Arabidopsis thaliana)"
     mRNA            complement(join(<11962. .12330,12423. .>12830))
                     /gene="OSJNBb0089A17.22"
     CDS             complement(join(11962. .12330,12423. .12830))
                     /gene="OSJNBb0089A17.22"
                     /codon_start=1
                     /product="putative transcriptional factor"
                     /protein_id="AAK31271.1"
                     /db_xref="GI:13569987"
                     /translation="MAMAKELQRTSSSSSSSAASTSSCSSAVTDAWSSPARPNAVAGG
                     KRKKHVVGRADEAAGGGAGEEEEEEAEAAAAGKSSAATKKRKRSSDGKHPVTRGVRMR
                     AWGRWVSEIREPRKSRTWLDTPPTAAMAAEAAHDVQAAAALAAAFTTSPSSSPSSSL
                     SADDVAPCVVHADADEQPAAAAKNDDDGSTTAPVAAAAAAAADEQQLPDLPDLLFD
                     IQDGPFGPAMWAPLADVDEVNAELRLEEPLLWDLGVTDA"
     repeat_region   complement(12126. .12162)
                     /rpt_family="(CGG)n"
     repeat_region   complement(12565. .12703)
                     /rpt_family="(GGA)n"
     repeat_region   12750. .12802
                     /rpt_family="(GGA)n"
     repeat_region   complement(13363. .13409)
                     /rpt_family="AT_rich"
     repeat_region   complement(13435. .13464)
                     /rpt_family="AT_rich"
     repeat_region   complement(15206. .15247)
                     /rpt_family="AT_rich"
     repeat_region   complement(15414. .15465)
                     /rpt_family="AT_rich"
     gene            complement(15582. .16007)
                     /gene="OSJNBb0089A17.21"
                     /note="predicted by glimmerR"
     mRNA            complement(<15582. .>16007)
                     /gene="OSJNBb0089A17.21"
     CDS             complement(15582. .16007)
                     /gene="OSJNBb0089A17.21"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAK31268.1"
                     /db_xref="GI:13569984"
                     /translation="MDYCRPSKPGVALAPWLSPGHWTVQTIVAATPRPSVQRTPSMR
                     AEMAAGTCAVEDASEHGGRVAGRCPERVRVHRIRARQWDDARRAAALRSIDRQYACID
                     RSPPWKTKAVDGLSVRIGIHSYDPKTHAFYLHKFVWAGA"
     repeat_region   complement(17572. .17613)
                     /rpt_family="(CGG)n"
     repeat_region   complement(17622. .17642)
                     /rpt_family="GC_rich"
     gene            17938. .21425
                     /gene="OSJNBb0089A17.1"
                     /note="similar to 8 amino 7 oxononanoate synthase
                     GB:JQ0512 GI:98180 (Bacillus sphaericus); EST AU094796
                     from this gene"
     mRNA            join(<17938. .18133,18368. .18480,18672. .18779,18871. .19109,
                     19352. .19439,19521. .19654,20178. .20266,20622. .20655,
                     20734. .20802,21120. .21425)
                     /gene="OSJNBb0089A17.1"
     CDS             join(17938. .18133,18368. .18480,18672. .18779,18871. .19109,
                     19352. .19439,19521. .19654,20178. .20266,20622. .20655,
                     20734. .20802,21120. .21243)
                     /gene="OSJNBb0089A17.1"
                     /codon_start=1
                     /product="putative 8-amino-7-oxononanoate synthase"
                     /protein_id="AAK31282.1"
                     /db_xref="GI:13569998"
                     /translation="MAQWDALVDAALARLASRSLLRATRPIALAPPPAPPFTPCPGP
                     WDRAAVEIRLDRATLHQWLAEGGEGSGQEEKVDEKLILFSGNDYMGLSSHPAIRHAAV
                     KAAEEYGMGPRGSALICGYTTYHKMVEESLAELKKEKDCLLCPTGFSANMAVMTALGN
                     ISSDIAAGKKPAEDERIAIFSDALNHASIIDGIRLVERQQNVSFVVKKCNHSHLELL
                     LLFSMDGGFAPLPDLVKLRKKYGFLLVIDDAHGTLVCGENSGQAAPELFECENKIDIS
                     GTLSKAAGCQGGFIACSTRWKRLIQSGRGRSFIFSTALPVPVVASVTVGSEKAALRAGR
                     DMLRSGFHVTPIRPPTVPPNSCKLRITLSASHSSDDIKRLVDALFPWLPKEKHDIQIYA
                     VASKL"
     repeat_region   complement(18024. .18046)
                     /rpt_family="GC_rich"
     repeat_region   24913. .24939
                     /rpt_family="(CAGTT)n"
     gene            complement(24934. .26046)
                     /gene="OSJNBb0089A17.20"
                     /note="predicted by fgenesh"
     mRNA            complement(join(<24934. .25129,25223. .>26046))
                     /gene="OSJNBb0089A17.20"
     CDS             complement(join(24934. .25129,25223. .26046))
                     /gene="OSJNBb0089A17.20"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAK31265.1"
                     /db_xref="GI:13569981"
                     /translation="MASSGAVALMMFHTWEERLFWRLVGSHGQQPGPMRWVIALWLKL
                     ESVGHHWFVRKVAVLPASVVLRFVSERLACLARLPHERGVAQGAEKRLAAAAAGDAD
                     PALRPLPCTNALLAEPVEGLAYFDARDEVMEGVSDVTRNVCRVIFDGVAAAVAAADD
                     DDDDAEAAAFLPRDVLDALDGTPPPPPPPPMYHQYHHRAVEMAPMLPPPPPYAELNPM
                     ASPWFPVQQQEQPPPPPPQPHQHGYIPLPEDYRSLFITFSRGYPIRQDDIINFFNSL
                     YGPCVRESVMVEKAAAGQLPVYGRVVLRCPSMIPVVLDGQQTAKYMIKGRHLWARIYVP
                     SSKPN"
     repeat_region   25332. .25357
                     /rpt_family="(CGG)n"
     repeat_region   25414. .25502
                     /rpt_family="(CGG)n"
     repeat_region   25541. .25594
                     /rpt_family="(CGG)n"
     repeat_region   complement(25734. .25822)
                     /rpt_family="GC_rich"
     gene            29995. .32244
                     /gene="OSJNBb0089A17.10"
                     /note="similar to quinone oxidoreductase GB:AAA23691
                     GI:145766 (Escherichia coli)"
     mRNA            join(<29995. .30039,30157. .30295,31110. .31336,31482. .31739
```

```
 Query Match          82.0%;  Score 16.4;  DB 8;  Length 132927;
 Best Local Similarity  94.4%;  Pred. No. 4.3e+02;
 Matches  17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 ATCTGAATCCGACCAACC 19
          ||||||||| ||||||||||
Db 105039 ATCTGAATCTGACCAACC 105056


RESULT 10
AC092272/c
LOCUS       AC092272               145668 bp    DNA     linear   PRI 24-OCT-2001
DEFINITION  Homo sapiens chromosome 5 clone CTC-339E17, complete sequence.
ACCESSION   AC092272
VERSION     AC092272.2  GI:16356877
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens.
```

```
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (bases 1 to 145668)
  AUTHORS       DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE         Direct Submission
  JOURNAL       Unpublished
REFERENCE       2  (bases 1 to 145668)
  AUTHORS       DOE Joint Genome Institute.
  TITLE         Direct Submission
  JOURNAL       Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
                Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE       3  (bases 1 to 145668)
  AUTHORS       DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE         Direct Submission
  JOURNAL       Submitted (24-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell
                Drive, Walnut Creek, CA 94598, USA
COMMENT         On Oct 24, 2001 this sequence version replaced gi:14589461.
                Draft Sequence Produced by DOE Joint Genome Institute
                www.jgi.doe.gov
                Finishing Completed at Stanford Human Genome Center
                www-shgc.stanford.edu
                Quality: Phrap Quality >=40 99.7% of Sequence;
                Estimated Total Number of Errors is 0.6.
                NOTE: This insert is not the entire sequence of the clone (entire
                sequence is 157kb). It is clipped at the overlap with AC008844. The
                number of bases overlapped is 47292.
FEATURES             Location/Qualifiers
     source          1..145668
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTC-339E17"
BASE COUNT     47790 a  28057 c  27632 g  42189 t
ORIGIN

  Query Match          82.0%;  Score 16.4;  DB 9;  Length 145668;
  Best Local Similarity   94.4%;  Pred. No. 4.3e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAAC 18
           ||||||||||| ||||||
Db   12146 GATCTGAATCCTACCAAC 12129


RESULT 11
AC019350/c
LOCUS       AC019350               155813 bp    DNA     linear   HTG 05-APR-2000
DEFINITION  Homo sapiens clone RP11-276018, WORKING DRAFT SEQUENCE, 9 unordered
            pieces.
ACCESSION   AC019350
VERSION     AC019350.3  GI:7417793
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 155813)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-276018
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 155813)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkly,L., BouKhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vasailler,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 5, 2000 this sequence version replaced gi:6715831.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5469
                Center clone name: 276_0_18
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 150710 bases at least Q40
                Consensus quality: 153191 bases at least Q30
                Consensus quality: 154025 bases at least Q20
                Insert size: 152000; agarose-fp
                Insert size: 155013; sum-of-contigs
                Quality coverage: 5.6 in Q20 bases; agarose-fp
                Quality coverage: 5.5 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     2535: contig of 2535 bp in length
            *    2536  2635: gap of     100 bp
            *    2636  5535: contig of 2900 bp in length
            *    5536  5635: gap of     100 bp
            *    5636  9094: contig of 3459 bp in length
            *    9095  9194: gap of     100 bp
            *    9195 15064: contig of 5870 bp in length
            *   15065 15164: gap of     100 bp
            *   15165 22781: contig of 7617 bp in length
            *   22782 22881: gap of     100 bp
            *   22882 39817: contig of 16936 bp in length
            *   39818 39917: gap of     100 bp
            *   39918 58842: contig of 18925 bp in length
            *   58843 58942: gap of     100 bp
            *   58943 102439: contig of 43497 bp in length
            *  102440 102539: gap of     100 bp
            *  102540 155813: contig of 53274 bp in length.
FEATURES             Location/Qualifiers
     source          1..155813
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-276018"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..2535
                     /note="assembly_fragment"
     misc_feature    2636..5535
                     /note="assembly_fragment"
     misc_feature    5636..9094
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    9195..15064
                     /note="assembly_fragment"
```

```
     misc_feature    15165. .22781
                     /note="assembly_fragment"
     misc_feature    22882. .39817
                     /note="assembly_fragment"
     misc_feature    39918. .58842
                     /note="assembly_fragment"
     misc_feature    58943. .102439
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    102540. .155813
                     /note="assembly_fragment"
BASE COUNT    51921 a  30751 c  29180 g  43159 t    802 others
ORIGIN

  Query Match             82.0%;  Score 16.4;  DB 2;  Length 155813;
  Best Local Similarity   94.4%;  Pred. No. 4.3e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GATCTGAATCCGACCAAC 18
          ||||||||||| ||||||
Db   1559 GATCTGAATCCTACCAAC 1542


RESULT 12
AC091914
LOCUS       AC091914               165237 bp    DNA     linear   PRI 27-FEB-2002
DEFINITION  Homo sapiens chromosome 5 clone RP11-221N1, complete sequence.
ACCESSION   AC091914
VERSION     AC091914.3  GI:18958640
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165237)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 165237)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 165237)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Feb 27, 2002 this sequence version replaced gi:14579738.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.5% of Sequence;
            Estimated Total Number of Errors is 0.5.
FEATURES             Location/Qualifiers
     source          1. .165237
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-221N1"
BASE COUNT    46852 a  30135 c  32136 g  56114 t
ORIGIN

  Query Match             82.0%;  Score 16.4;  DB 9;  Length 165237;
  Best Local Similarity   94.4%;  Pred. No. 4.3e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GATCTGAATCCGACCAAC 18
          ||||||||||| ||||||
Db  37737 GATCTGAATCCTACCAAC 37754


RESULT 13
AC114308
LOCUS       AC114308               177246 bp    DNA     linear   HTG 07-MAR-2002
DEFINITION  Homo sapiens chromosome 5 clone RP11-197F21, WORKING DRAFT
            SEQUENCE, 5 unordered pieces.
ACCESSION   AC114308
VERSION     AC114308.1  GI:19224957
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 177246)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 177246)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 482154
            Center clone name: RPCI-11_197F21
            -------
            Summary Statistics
            Consensus quality: 171433 bases at least Q40
            Consensus quality: 173924 bases at least Q30
            Consensus quality: 174945 bases at least Q20
            Estimated insert size: 175000; agarose-fp estimation
            Estimated insert size: 176846; sum-of-contigs estimation
            Quality coverage:  9.88 in Q20 bases; agarose-fp estimation
            Quality coverage:  9.77 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1      1270: contig of 1270 bp in length
            *    1271      1370: gap of unknown length
            *    1371     12248: contig of 10878 bp in length
            *   12249     12348: gap of unknown length
            *   12349     33985: contig of 21637 bp in length
            *   33986     34085: gap of unknown length
            *   34086     79299: contig of 45214 bp in length
            *   79300     79399: gap of unknown length
            *   79400    177246: contig of 97847 bp in length.
FEATURES             Location/Qualifiers
     source          1. .177246
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-197F21"
                     /clone_lib="RPCI human BAC library 11"
BASE COUNT    54617 a  32530 c  33575 g  56124 t    400 others
ORIGIN

  Query Match             82.0%;  Score 16.4;  DB 2;  Length 177246;
  Best Local Similarity   94.4%;  Pred. No. 4.3e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GATCTGAATCCGACCAAC 18
          ||||||||||| ||||||
Db  84991 GATCTGAATCCTACCAAC 85008
```

```
RESULT 14
AC018248/c
LOCUS       AC018248               9359 bp    DNA     linear   HTG 09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC018248
VERSION     AC018248.1  GI:6552943
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 9359)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10214038 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..9359
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     2737 a   2174 c   1909 g   2539 t
ORIGIN

  Query Match           80.0%;  Score 16;  DB 2;  Length 9359;
  Best Local Similarity  100.0%;  Pred. No. 7.4e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ATCTGAATCCGACCAA 17
           ||||||||||||||||
Db    7973 ATCTGAATCCGACCAA 7958


RESULT 15
AC012160/c
LOCUS       AC012160               172069 bp    DNA     linear   INV 13-DEC-2001
DEFINITION  Drosophila melanogaster, chromosome X, region 15E-15E, BAC clone
            BAC06G02, complete sequence.
ACCESSION   AC012160
VERSION     AC012160.6  GI:17646875
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 172069)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dragnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome X, region 15E-15E
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 172069)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blaze},R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Dec 13, 2001 this sequence version replaced gi:6466928.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1..172069
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="X"
                     /map="15E-15E"
                     /clone="BACR06G02 (DI110)"
                     /clone_lib="BPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
BASE COUNT    48971 a  35590 c  36029 g  51479 t
ORIGIN

  Query Match           80.0%;  Score 16;  DB 3;  Length 172069;
  Best Local Similarity  100.0%;  Pred. No. 7.2e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ATCTGAATCCGACCAA 17
           ||||||||||||||||
Db  162241 ATCTGAATCCGACCAA 162226


Search completed: May 12, 2003, 01:51:34
Job time : 391 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 00:45:59 ; Search time 35.2143 Seconds
(without alignments)
1279.025 Million cell updates/sec

Title: US-09-980-277-2
Perfect score: 20
Sequence: 1 gatctgaatccgaccaaccg 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	20	100.0	20	22	AAI70401	Schistosoma manson
	2	20	100.0	120	22	AAI70400	Schistosoma manson
	3	20	100.0	6494	21	AAA46170	GFP-Sm1-7 fusion p
	4	16.8	84.0	467	18	AAT83836	DNA encoding a Sta
	5	16.8	84.0	22188	23	ABL28616	Drosophila melanog
c	6	16	80.0	2889	23	ABL16704	Drosophila melanog
	7	15.8	79.0	20	19	AAV51591	Zea mays genome fo
	8	15.8	79.0	530	22	AAR11862	Human brain expres
c	9	15.8	79.0	1005	23	AAS55514	Streptococcus pneu
c	10	15.8	79.0	1173	24	ABQ53190	Oligonucleotide fo
	11	15.8	79.0	1173	24	ABQ53191	Oligonucleotide fo
c	12	15.8	79.0	1210	24	ABQ52080	Oligonucleotide fo
	13	15.8	79.0	1210	24	ABQ52081	Oligonucleotide fo
	14	15.4	77.0	22	24	ABA90718	Lactococcus lactis
	15	15.4	77.0	723	23	ABL21263	Drosophila melanog
	16	15.4	77.0	724	21	AAF11860	Aspergillus oryzae
	17	15.4	77.0	2259	23	ABL21262	Drosophila melanog
	18	15.4	77.0	2733	23	ABL21266	Drosophila melanog
	19	15.4	77.0	13518	23	ABL03288	Drosophila melanog
	20	15.2	76.0	504	24	ABL56299	KGMMV virus coat p
c	21	15.2	76.0	534	22	AAX92799	Human cDNA 3'-end
c	22	15.2	76.0	1153	24	ABQ18462	Oligonucleotide fo
	23	15.2	76.0	1153	24	ABQ18463	Oligonucleotide fo
	24	15.2	76.0	1531	20	AAX80901	Cuphea viscosissim
	25	15.2	76.0	1620	21	AAZ52445	BTEM clone 2495790
	26	15.2	76.0	2091	21	AAC77593	Human ORFX ORF3148
	27	15.2	76.0	2460	21	AAA78413	Human secreted pro
	28	15.2	76.0	2627	22	AAR34378	Human full-length
c	29	15.2	76.0	7914	22	AAF84648	Plasmid DLZ7 encod
c	30	15.2	76.0	7944	22	AAF84647	Plasmid DLZ6 encod
	31	15.2	76.0	19929	22	AAL07279	Human reproductive
	32	15.2	76.0	19929	23	ABL98825	Human testicular a
	33	15.2	76.0	20907	22	AAL07278	Human reproductive
	34	15.2	76.0	20907	23	ABL98824	Human testicular a
c	35	15.2	76.0	31271	23	ABL26680	Drosophila melanog
	36	15.2	76.0	534720	19	AAV30458	Rhizobium species
	37	15.2	76.0	536165	19	AAV30459	Rhizobium species
c	38	15	75.0	1345	24	ABQ15648	Oligonucleotide fo
	39	15	75.0	1345	24	ABQ15649	Oligonucleotide fo
	40	15	75.0	3562	23	ABL06749	Drosophila melanog
	41	15	75.0	7143	23	ABL06748	Drosophila melanog
c	42	15	75.0	8186	24	ABN80090	Human chemically m
	43	14.8	74.0	319	21	AAC30386	Human secreted pro
	44	14.8	74.0	437	24	ABL80093	Human ovarian canc
c	45	14.8	74.0	451	22	ABA12238	Human nervous syst

ALIGNMENTS

RESULT 1
AAI70401
ID AAI70401 standard; DNA; 20 BP.
XX
AC AAI70401;
XX
DT 07-JAN-2002 (first entry)
XX
DE Schistosoma mansoni repeater unit 5' PCR primer.
XX
KW Schistosomiasis; infection; diagnosis; PCR primer; ss.
XX
OS Schistosoma mansoni.
XX
PN WO200175148-A1.
XX
PD 11-OCT-2001.
XX
PF 04-APR-2001; 2001WO-BR00035.
XX
PR 04-APR-2000; 2000BR-0001536.
XX
PA (FIOC-) FIOCRUZ FUNDACAO CRUZ OSWALDO.
XX
PI Teles Rabello AL, Dias Neto E, Pontes LA;
XX
DR WPI; 2001-648561/74.
XX
PT Diagnosing infection by parasites of the Schistosoma sp. in a
PT biological sample especially useful in cases of low infection
PT intensity, comprises detecting a specific region of the DNA of
PT Schistosoma by polymerase chain reaction -

```
XX
PS   Claim 4; Page 13; 34pp; English.
XX
CC   The present sequence is that of a claimed oligonucleotide primer,
CC   derived from the 5' region of a highly repeated short DNA sequence
CC   (see AAI70400) of the Schistosoma mansoni genome.  This 5' primer,
CC   and the 3' primer given in AAI70402, are used in the method of the
CC   invention for diagnosing Schistosoma infection.  The method involves
CC   collection of the sample to be examined, extraction of Schistosoma
CC   sp. DNA, amplification by PCR, separation of PCR products by
CC   electrophoresis, and detection e.g. by colouring with silver salts.
CC   A kit for diagnosing infection, which includes the primers, is
CC   claimed.  The method is useful for detecting Schistosoma sp.
CC   parasites by detecting the parasite's DNA in a biological sample.
CC   It is especially useful in cases of low infection intensity for
CC   which parasitological stool tests demonstrate little sensitivity.
XX
SQ   Sequence 20 BP; 6 A; 7 C; 4 G; 3 T; 0 other;

  Query Match             100.0%;  Score 20;  DB 22;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 1.1;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACCG 20
           ||||||||||||||||||||
Db       1 GATCTGAATCCGACCAACCG 20


RESULT 2
AAI70400
ID   AAI70400 standard; DNA; 120 BP.
XX
AC   AAI70400;
XX
DT   07-JAN-2002  (first entry)
XX
DE   Schistosoma mansoni repeater unit DNA.
XX
KW   Schistosomiasis; infection; diagnosis; ds.
XX
OS   Schistosoma mansoni.
XX
FH   Key             Location/Qualifiers
FT   primer_bind     complement (1..19)
FT                   /tag= a
FT                   /note= "primer of AAI70401"
FT   primer_bind     91..110
FT                   /tag= b
FT                   /note= "primer of AAI70402"
XX
PN   WO200175148-A1.
XX
PD   11-OCT-2001.
XX
PF   04-APR-2001; 2001WO-BR00035.
XX
PR   04-APR-2000; 2000BR-0001536.
XX
PA   (FIOC-) FIOCRUZ FUNDACAO CRUZ OSWALDO.
XX
PI   Teles Rabello AL,  Dias Neto E,  Pontes LA;
XX
DR   WPI; 2001-648561/74.
XX
PT   Diagnosing infection by parasites of the Schistosoma sp. in a
PT   biological sample especially useful in cases of low infection
PT   intensity, comprises detecting a specific region of the DNA of
PT   Schistosoma by polymerase chain reaction -
XX
PS   Claim 1; Page 22; 34pp; English.
XX
CC   The present sequence is that of a highly repeated short DNA
CC   sequence in the genome of Schistosoma mansoni.  PCR primers
CC   (see AAI70401 and AAI70402) that flank this sequence are used in
CC   the method of the invention for diagnosing Schistosoma infection.
CC   The method involves collection of the sample to be examined,
CC   extraction of Schistosoma sp. DNA, amplification by PCR, separation
CC   of PCR products by electrophoresis, and detection e.g. by colouring
CC   with silver salts.  A kit for diagnosing infection, which includes
CC   the primers, is claimed.  The method is useful for detecting
CC   Schistosoma sp. parasites by detecting the DNA of the parasite in a
CC   biological sample.  It is especially useful in cases of low
CC   infection intensity for which parasitological stool tests
CC   demonstrate little sensitivity.
XX
SQ   Sequence 120 BP; 36 A; 20 C; 26 G; 38 T; 0 other;

  Query Match             100.0%;  Score 20;  DB 22;  Length 120;
  Best Local Similarity   100.0%;  Pred. No. 1.3;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACCG 20
           ||||||||||||||||||||
Db       1 GATCTGAATCCGACCAACCG 20


RESULT 3
AAA46170
ID   AAA46170 standard; DNA; 6494 BP.
XX
AC   AAA46170;
XX
DT   27-SEP-2000  (first entry)
XX
DE   GFP-Sm1-7 fusion protein construct.
XX
KW   GFP-Sm1-7 fusion construct; circular; green fluorescent protein;
KW   glutathione S-transferase; eukaryotic diploid multicellular parasite;
KW   universal graft; transgenic eukaryotic parasite; acquired deficiency;
KW   genetic deficiency; hormone deficiency; metabolic deficiency;
KW   haematological deficiency; immunological deficiency; immunotherapy;
KW   anti-microbial therapy; anti-cancer therapy; drug addiction;
KW   poisoning condition; geriatric condition; ds.
XX
OS   Synthetic.
XX
PN   WO2000032804-A1.
XX
PD   08-JUN-2000.
XX
PF   01-DEC-1999;  99WO-IL00651.
XX
PR   01-DEC-1998;  98US-0201850.
XX
PA   (YISS ) YISSUM RES & DEV CO.
XX
PI   Hamburger J,  Laban A;
XX
DR   WPI; 2000-412348/35.
XX
PT   Eukaryotic diploid multicellular parasite useful as universal grafts
PT   for in vivo delivery of beneficial gene products in humans and animals
PT   involves transformation with a transgene  -
XX
PS   Example 2; Fig 6; 90pp; English.
XX
CC   This sequence represents a GFP-Sm1-7 fusion construct contained within a
CC   recombinant vector. This sequence contains the green fluorescent protein
CC   (GFP) coding sequence from Aequorea victoria, and the
CC   SM1-7 coding sequence from Schistosoma, along with
CC   promoter sequences. The invention relates to a eukaryotic diploid
CC   multicellular parasite transformed with a transgene. Transgenic
CC   eukaryotic parasites are useful as universal grafts for in vivo delivery
CC   of beneficial gene product in humans and animals. The parasites can
CC   particularly be used for restoration of deficiencies whether acquired or
CC   genetic, such as hormone deficiencies, metabolic deficiencies,
```

```
CC   haematological deficiencies, immunological deficiencies, immunotherapy,
CC   anti-microbial therapy, anti-cancer therapy. They can also be used for
CC   treatment of drug addiction, of poisoning conditions, and for
CC   amelioration of geriatric conditions. Treatment of humans with in vivo
CC   transgenes are universally compatible, readily available and inexpensive.
CC   Genotypic alterations of the patients is avoided, reducing risks of
CC   mutagenesis and malignant transformation.
XX
SQ   Sequence 6494 BP; 1754 A; 1373 C; 1481 G; 1886 T; 0 other;

  Query Match             100.0%;  Score 20;  DB 21;  Length 6494;
  Best Local Similarity   100.0%;  Pred. No. 1.6;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACCG 20
           ||||||||||||||||||||
Db    3055 GATCTGAATCCGACCAACCG 3074


RESULT 4
AAT83836
ID   AAT83836 standard; DNA; 467 BP.
XX
AC   AAT83836;
XX
DT   20-AUG-1998  (first entry)
XX
DE   DNA encoding a Staphylococcus aureus protein of unknown function.
XX
KW   Staphylococcus aureus protein; ribozyme; antisense sequence; control;
KW   Staphylococcal gene; regulatory element; bacterial gene expression;
KW   vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
KW   toxic shock syndrome; ss.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   CDS             144..419
FT                   /*tag= a
XX
PN   WO9730070-A1.
XX
PD   21-AUG-1997.
XX
PF   19-FEB-1997;   97WO-US02318.
XX
PR   20-FEB-1996;   96US-0011888.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Black MT,  Burnham MK,  Hodgson JE,  Knowles DJC,  Nicholas RO;
PI   Pratt JM,  Reichard RW,  Rosenberg M,  Ward JM;
XX
DR   WPI; 1997-424969/39.
DR   P-PSDB; RAW27874.
XX
PT   Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
PT   to isolate antimicrobial compounds, and in vaccines against S.
PT   aureus infection
XX
PS   Claim 9; Page 702; 989pp; English.
XX
CC   The present sequence encodes a Staphylococcus aureus protein of
CC   unknown function. The present sequence was isolated from a
CC   library of clones of S. aureus WCUH 29 in Escherichia coli. The DNA
CC   sequence can be used in the construction of ribozymes and antisense
CC   sequences to control the expression of Staphylococcal genes. The DNA
CC   sequence is also useful as a source of regulatory elements for the
CC   control of bacterial gene expression. The encoded protein may be used
CC   to produce vaccines to enable a host to produce specific antibodies
CC   with antibacterial action. These vaccines and antibodies would protect
CC   a host against invasion by S. aureus, and conditions relating to
CC   Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled
CC   skin syndrome, and toxic shock syndrome.
XX
SQ   Sequence 467 BP; 192 A; 83 C; 86 G; 99 T; 7 other;

  Query Match             84.0%;  Score 16.8;  DB 18;  Length 467;
  Best Local Similarity   90.0%;  Pred. No. 56;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACCG 20
           ||| ||||| ||||||||||
Db      34 GATTTGAATGCGACCAACCG 53


RESULT 5
ABL28616
ID   ABL28616 standard; DNA; 22188 BP.
XX
AC   ABL28616;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 37321.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 37321; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (AB857737-ABB73072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 22188 BP; 6664 A; 5376 C; 4559 G; 5589 T; 0 other;

  Query Match             84.0%;  Score 16.8;  DB 23;  Length 22188;
  Best Local Similarity   90.0%;  Pred. No. 71;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACCG 20
           || |||||||||| |||||
Db    1843 GAACTGAATCCGACAAACCG 1862


RESULT 6
```

```
ABL16704/c
ID   ABL16704 standard; DNA; 2889 BP.
XX
AC   ABL16704;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 1585.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 1585; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL19511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2889 BP; 863 A; 695 C; 561 G; 770 T; 0 other;

  Query Match             80.0%;  Score 16;  DB 23;  Length 2889;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ATCTGAATCCGACCAA 17
           ||||||||||||||||
Db     971 ATCTGAATCCGACCAA 956


RESULT 7
AAV51591
ID   AAV51591 standard; DNA; 20 BP.
XX
AC   AAV51591;
XX
DT   02-FEB-1999  (first entry)
XX
DE   Zea mays genome forward PCR primer #191.
XX
KW   Polymorphic marker; allele-specific; probe; amplification; PCR primer;
KW   hybridisation; plant; hybrid certification; genetic contribution;
KW   progeny; back-cross; hybrid; ancestry; corn; ss.
XX
OS   Synthetic.
XX
OS   Zea mays.
XX
PN   WO9824796-A1.
XX
PD   11-JUN-1998.
XX
PF   01-DEC-1997;   97WO-US21782.
XX
PR   07-MAR-1997;   97US-0813507.
PR   02-DEC-1996;   96US-0032069.
XX
PA   (AFFY-) AFFYMETRIX INC.
XX
PI   Landry BS,  Lemieux B,  Murigneux A,  Sapolsky RJ;
XX
DR   WPI; 1998-333252/29.
XX
PT   Brassica species allele-specific oligonucleotide probes and primers
PT   - useful for plant breeding
XX
PS   Example 1; Page Page 53; 65pp; English.
XX
CC   AAV51401-V51704 are forward PCR primers used to amplify fragments of the
CC   Zea mays genome in order to detect polymorphic markers. Such markers can
CC   be used in the construction of allele-specific primers and probes for
CC   amplification or hybridisation, e.g. to determine common or disparate
CC   ancestry between 2 or more plants, to monitor the genetic contribution
CC   of an ancestral plant, to trace the progeny of proprietary plants, in
CC   certification of a hybrid plant or to identify the progeny of a
CC   back-crossed plant with an ancestral plant.
XX
SQ   Sequence 20 BP; 5 A; 8 C; 3 G; 4 T; 0 other;

  Query Match             79.0%;  Score 15.8;  DB 19;  Length 20;
  Best Local Similarity   89.5%;  Pred. No. 1.5e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACC 19
           |||||||||| |||| |||
Db       1 GATCTGAATCTGACCCACC 19


RESULT 8
AAK11862
ID   AAK11862 standard; DNA; 530 BP.
XX
AC   AAK11862;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human brain expressed single exon probe SEQ ID NO: 11853.
XX
KW   Human; brain expressed exon; gene expression analysis; probe;
KW   microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW   epilepsy; cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157275-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00667.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
```

```
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483446/52.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   brains -
XX
PS   Example 4; SEQ ID NO: 11853; 650pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   brain. They can be used to measure gene expression in brain cell samples,
CC   which may enable the diagnosis and improved treatment of nervous system
CC   diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC   epilepsy and cancers. The present sequence is one of the probes of the
CC   invention.
XX
SQ   Sequence 530 BP; 151 A; 132 C; 59 G; 188 T; 0 other;

  Query Match             79.0%;  Score 15.8;  DB 22;  Length 530;
  Best Local Similarity   89.5%;  Pred. No. 1.8e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 ATCTGAATCCGACCAACCG 20
          || ||||||||||||||| |
Db      2 ATGTGAATCCGACCAACTG 20


RESULT 9
AAS55514/c
ID   AAS55514 standard; DNA; 1005 BP.
XX
AC   AAS55514;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Streptococcus pneumoniae DNA for cellular proliferation protein #85.
XX
KW   Antisense; ds; prokaryotic cellular proliferation gene;
KW   antibiotic; antibacterial; drug design.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   P-PSDB; AAU37655.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Claim 27; Seq ID No 9151; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence encodes an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1005 BP; 277 A; 175 C; 255 G; 298 T; 0 other;

  Query Match             79.0%;  Score 15.8;  DB 23;  Length 1005;
  Best Local Similarity   89.5%;  Pred. No. 1.9e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 ATCTGAATCCGACCAACCG 20
          ||||||||||||||| ||| |
Db    596 ATCTGAATCCGACGAACTG 578


RESULT 10
ABQ53190/c
ID   ABQ53190 standard; DNA; 1173 BP.
XX
AC   ABQ53190;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 39781.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
```

```
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 1173 BP; 164 A; 142 C; 451 G; 416 T; 0 other;

  Query Match             79.0%;  Score 15.8;  DB 24;  Length 1173;
  Best Local Similarity   89.5%;  Pred. No. 1.9e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 ATCTGAATCCGACCAACCG 20
           | || ||||||||||||||
Db     920 ACCTAAATCCGACCAACCG 902


RESULT 11
ABQ53191
ID   ABQ53191 standard; DNA; 1173 BP.
XX
AC   ABQ53191;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 39782.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 1173 BP; 416 A; 451 C; 142 G; 164 T; 0 other;

  Query Match             79.0%;  Score 15.8;  DB 24;  Length 1173;
  Best Local Similarity   89.5%;  Pred. No. 1.9e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 ATCTGAATCCGACCAACCG 20
           | || ||||||||||||||
Db     254 ACCTAAATCCGACCAACCG 272


RESULT 12
ABQ52080/c
ID   ABQ52080 standard; DNA; 1210 BP.
XX
AC   ABQ52080;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 38671.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
```

CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 1210 BP; 139 A; 153 C; 399 G; 519 T; 0 other;

Query Match 79.0%; Score 15.8; DB 24; Length 1210;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ATCTGAATCCGACCAACCG 20
||| ||| ||||||||||||
Db 506 ATCCGAACCCGACCAACCG 488

RESULT 13
ABQ52081
ID ABQ52081 standard; DNA; 1210 BP.
XX
AC ABQ52081;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 38672.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (1) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 1210 BP; 519 A; 399 C; 153 G; 139 T; 0 other;

Query Match 79.0%; Score 15.8; DB 24; Length 1210;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ATCTGAATCCGACCAACCG 20
||| ||| ||||||||||||
Db 705 ATCCGAACCCGACCAACCG 723

RESULT 14
ABA90718
ID ABA90718 standard; DNA; 22 BP.
XX
AC ABA90718;
XX
DT 16-MAY-2002 (first entry)
XX
DE Lactococcus lactis oligonucleotide #197 used in Long Range PCR.
XX
KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese; ss.
XX
OS Lactococcus lactis IL1403.
XX
PN FR2807446-A1.
XX
PD 12-OCT-2001.
XX
PF 11-APR-2000; 2000FR-0004630.
XX
PR 11-APR-2000; 2000FR-0004630.
XX
PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX
PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;
XX
DR WPI; 2002-043418/06.
XX
PT New nucleotide sequence useful in the identification or Lactococcus
PT lactis and related species -
XX
PS Example 1; SEQ ID No 2520; 2504pp; French.
XX
CC The present invention is related to a Lactococcus lactis nucleotide
CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
CC nucleic acid sequence is useful in the detection and/or amplification of
CC nucleic acid sequence, particularly to identify Lactococcus lactis or
CC related species. The proteins of the invention are useful for the
CC biosynthesis or biodegradation of a composition of interest. The
CC invention helps research in lactic bacteria, particularly useful in the
CC production of yogurt and cheese. The present sequence is an
CC oligonucleotide used in an example from the invention.
CC Note: The sequence data for this patent is based on equivalent patent
CC WO200177334 (published 18-OCT-2001) which is available in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 22 BP; 7 A; 7 C; 5 G; 3 T; 0 other;

Query Match 77.0%; Score 15.4; DB 24; Length 22;
Best Local Similarity 94.1%; Pred. No. 2.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 CTGAATCCGACCAACCG 20
||||||||||||||| ||
Db 5 CTGAATCCGACCAAGCG 21

```
RESULT 15
ABL21263
ID   ABL21263 standard; DNA; 723 BP.
XX
AC   ABL21263;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 15262.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 15262; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 723 BP; 163 A; 195 C; 193 G; 172 T; 0 other;

  Query Match           77.0%;  Score 15.4;  DB 23;  Length 723;
  Best Local Similarity   94.1%;  Pred. No. 2.9e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 TCTGAATCCGACCAACC 19
           ||||||| ||||||||
Db     303 TCTGAATCTGACCAACC 319


Search completed: May 12, 2003, 01:14:56
Job time : 40.2143 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 01:08:23 ; Search time 8.78571 Seconds
(without alignments)
698.126 Million cell updates/sec

Title: US-09-980-277-2
Perfect score: 20
Sequence: 1 gatctgaatccgaccaaccg 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	15.8	79.0	3671	4	US-08-932-787B-5	Sequence 5, Appli
2	15.8	79.0	3671	4	US-08-932-012C-5	Sequence 5, Appli
3	15.8	79.0	3671	4	US-08-888-818C-5	Sequence 5, Appli
4	15.2	76.0	1931	2	US-08-948-176-24	Sequence 24, Appl
c 5	15	75.0	7096	4	US-09-221-017B-373	Sequence 373, App
6	14.8	74.0	2794	1	US-08-435-925C-1	Sequence 1, Appli
7	14.8	74.0	4843	3	US-08-986-485-1	Sequence 1, Appli
8	14.2	71.0	502	4	US-09-398-992-497	Sequence 497, App
9	14.2	71.0	1080	4	US-09-098-219b-1	Sequence 1, Appli
10	14.2	71.0	2916	4	US-08-976-259-22	Sequence 22, Appl
11	14.2	71.0	3671	1	US-08-176-620A-5	Sequence 5, Appli
12	14.2	71.0	3671	1	US-08-463-862-5	Sequence 5, Appli
13	14.2	71.0	3671	2	US-08-461-985-5	Sequence 5, Appli
14	14.2	71.0	3671	2	US-08-458-887-5	Sequence 5, Appli
c 15	14.2	71.0	3677	2	US-08-390-888A-2	Sequence 2, Appli
c 16	14.2	71.0	6496	4	US-09-221-017B-543	Sequence 543, App
17	14.2	71.0	10846	4	US-09-098-219b-5	Sequence 5, Appli
18	14.2	71.0	10900	4	US-09-098-219b-6	Sequence 6, Appli
19	13.8	69.0	555	4	US-08-479-089A-4	Sequence 4, Appli
c 20	13.8	69.0	691	4	US-08-961-527-381	Sequence 381, App
c 21	13.8	69.0	1003	4	US-08-961-527-378	Sequence 378, App
22	13.8	69.0	1168	4	US-08-961-527-362	Sequence 362, App
23	13.8	69.0	1197	4	US-09-134-001C-2532	Sequence 2532, Ap
24	13.8	69.0	1949	4	US-08-961-527-289	Sequence 289, App
c 25	13.8	69.0	2335	4	US-08-961-527-275	Sequence 275, App
c 26	13.8	69.0	2393	4	US-08-961-527-90	Sequence 90, Appl
27	13.8	69.0	2526	1	US-07-912-952-1	Sequence 1, Appli
28	13.8	69.0	3763	4	US-08-961-527-186	Sequence 186, App
c 29	13.8	69.0	4137	4	US-09-221-017B-329	Sequence 329, App
30	13.8	69.0	4266	4	US-09-651-011A-3	Sequence 3, Appli
c 31	13.8	69.0	6492	4	US-08-961-527-188	Sequence 188, App
c 32	13.8	69.0	6516	4	US-08-961-527-105	Sequence 105, App
33	13.6	68.0	30	4	US-09-545-244A-3	Sequence 3, Appli
34	13.6	68.0	38	3	US-09-109-063-16	Sequence 16, Appl
c 35	13.6	68.0	41	3	US-09-109-063-17	Sequence 17, Appl
36	13.6	68.0	218	3	US-09-035-648-14	Sequence 14, Appl
37	13.6	68.0	218	4	US-09-001-951-14	Sequence 14, Appl
38	13.6	68.0	218	4	US-08-818-829-14	Sequence 14, Appl
c 39	13.6	68.0	372	3	US-08-908-643C-4	Sequence 4, Appli
40	13.6	68.0	827	5	PCT-US95-06406A-20	Sequence 20, Appl
41	13.6	68.0	993	3	US-09-109-063-2	Sequence 2, Appli
c 42	13.6	68.0	1176	3	US-09-689-421-18	Sequence 18, Appl
c 43	13.6	68.0	1176	4	US-09-389-528-18	Sequence 18, Appl
c 44	13.6	68.0	1176	4	US-09-181-827A-18	Sequence 18, Appl
c 45	13.6	68.0	1347	3	US/08/622	INFORMATION FOR

ALIGNMENTS

RESULT 1
US-08-932-787B-5
; Sequence 5, Application US/08932787B
; Patent No. 6277953
; GENERAL INFORMATION:
; APPLICANT: Boulton et al.
; TITLE OF INVENTION: ANTIBODIES DIRECTED TOWARD EXTRACELLULAR SIGNAL-RELATED
; TITLE OF INVENTION: KINASES
; FILE REFERENCE: RBS 430-A-1
; CURRENT APPLICATION NUMBER: US/08/932,787B
; CURRENT FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 08/469,547
; PRIOR FILING DATE: 1995-06-06
; PRIOR APPLICATION NUMBER: 08/178,488
; PRIOR FILING DATE: 1994-01-07
; PRIOR APPLICATION NUMBER: 07/701,544
; PRIOR FILING DATE: 1991-05-16
; PRIOR APPLICATION NUMBER: 07/532,004
; PRIOR FILING DATE: 1990-06-01
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 3671
; TYPE: DNA
; ORGANISM: RAT
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (303)...(2018)
; OTHER INFORMATION: ERK3 cDNA
US-08-932-787B-5

Query Match 79.0%; Score 15.8; DB 4; Length 3671;
Best Local Similarity 89.5%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GATCTGAATCCGACCAACC 19
|||||||| ||| ||||||
Db 843 GATCTGAAGCCGGCCAACC 861

RESULT 2
US-08-932-012C-5
; Sequence 5, Application US/08932012C
; Patent No. 6297035
; GENERAL INFORMATION:
; APPLICANT: Boulton et al.
; TITLE OF INVENTION: A FAMILY OF MAP2 PROTEIN KINASES
; FILE REFERENCE: RBS 430-T-1
; CURRENT APPLICATION NUMBER: US/08/932,012C
; CURRENT FILING DATE: 1997-09-17

```
;  PRIOR APPLICATION NUMBER: 08/462,874
;  PRIOR FILING DATE: 1995-06-05
;  PRIOR APPLICATION NUMBER: 08/178,488
;  PRIOR FILING DATE: 1994-01-07
;  PRIOR APPLICATION NUMBER: 07/701,544
;  PRIOR FILING DATE: 1991-05-16
;  PRIOR APPLICATION NUMBER: 07/532,004
;  PRIOR FILING DATE: 1990-06-01
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 3671
;   TYPE: DNA
;   ORGANISM: RAT
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (303)...(2018)
;   OTHER INFORMATION: ERK3 cDNA
US-08-932-012C-5

  Query Match             79.0%;  Score 15.8;  DB 4;  Length 3671;
  Best Local Similarity   89.5%;  Pred. No. 19;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACC 19
           ||||||||| ||| |||||||
Db     843 GATCTGAAGCCGGCCAACC 861


RESULT 3
US-08-888-818C-5
; Sequence 5, Application US/08888818C
; Patent No. 6303358
; GENERAL INFORMATION:
;  APPLICANT:  Boulton et al.
;  TITLE OF INVENTION:  A FAMILY OF MAP2 PROTEIN KINASES
;  FILE REFERENCE:  REG 430-V-1
;  CURRENT APPLICATION NUMBER: US/08/888,818C
;  CURRENT FILING DATE:  1997-07-07
;  PRIOR APPLICATION NUMBER:  08/478,985
;  PRIOR FILING DATE:  1995-06-07
;  PRIOR APPLICATION NUMBER:  08/178,488
;  PRIOR FILING DATE:  1994-01-07
;  PRIOR APPLICATION NUMBER:  07/701,544
;  PRIOR FILING DATE:  1991-05-16
;  PRIOR APPLICATION NUMBER:  07/532,004
;  PRIOR FILING DATE:  1990-06-01
;  NUMBER OF SEQ ID NOS:  21
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 3671
;   TYPE: DNA
;   ORGANISM: RAT
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (303)...(2018)
;   OTHER INFORMATION: ERK3 cDNA
US-08-888-818C-5

  Query Match             79.0%;  Score 15.8;  DB 4;  Length 3671;
  Best Local Similarity   89.5%;  Pred. No. 19;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACC 19
           ||||||||| ||| |||||||
Db     843 GATCTGAAGCCGGCCAACC 861


RESULT 4
US-08-948-176-24
; Sequence 24, Application US/08948176
; Patent No. 5945585
;  GENERAL INFORMATION:
;    APPLICANT:  HITZ, WILLIAM D.
;    APPLICANT:  YADAV, NARENDRA S.
;    TITLE OF INVENTION:  ACYL-ACP THIOESTERASES GENES
;    TITLE OF INVENTION:  AND THEIR USE IN ALTERING PLANT
;    TITLE OF INVENTION:  OIL COMPOSITION
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.50 INCH
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WORD FOR WINDOWS 95
;      SOFTWARE:  MICROSOFT WORD VERSION 7.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/948,176
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/631,264
;      FILING DATE:  DECEMBER 20, 1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CHRISPENSEN, LYNNE M.
;      REGISTRATION NUMBER:  30,971
;      REFERENCE/DOCKET NUMBER:  CR-8926-C
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-992-5481
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1531 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  80..1192
US-08-948-176-24

  Query Match             76.0%;  Score 15.2;  DB 2;  Length 1531;
  Best Local Similarity   85.0%;  Pred. No. 35;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACCG 20
           ||||| |||| |||||||
Db    1062 GATCTGCATCCACCCAACCG 1081


RESULT 5
US-09-221-017B-373/c
; Sequence 373, Application US/09221017B
; Patent No. 6444799
;  GENERAL INFORMATION:
;    APPLICANT:  Ross, Bruce C.
;    TITLE OF INVENTION:  P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
;    NUMBER OF SEQUENCES:  1120
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 PAGE MILL ROAD
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/221,017B
;      FILING DATE:  23-DEC-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1182
;      FILING DATE:  31-DEC-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1546
;      FILING DATE:  30-JAN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP2911
;      FILING DATE:  09-APR-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU98/01023
;      FILING DATE:  10-DEC-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monroy, Gladys H
;      REGISTRATION NUMBER:  32,430
;      REFERENCE/DOCKET NUMBER:  27340-20021.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-813-5600
;      TELEFAX:  650-494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  373:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7096 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  UNKNOWN
;    ORIGINAL SOURCE:
;      ORGANISM:  PORYPHYROMONAS GINGIVALIS
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1...7096
US-09-221-017B-373

  Query Match             75.0%;  Score 15;  DB 4;  Length 7096;
  Best Local Similarity   100.0%;  Pred. No. 55;
  Matches  15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 TGAATCCGACCAACC 19
           |||||||||||||||
Db    4931 TGAATCCGACCAACC 4917


RESULT 6
US-08-435-925C-1
; Sequence 1, Application US/08435925C
; Patent No. 5646025
;    GENERAL INFORMATION:
;      APPLICANT:  Moyer, Donna
;      TITLE OF INVENTION:  SCYTALIDUM CATALASE GENE
;      NUMBER OF SEQUENCES:  18
;      CORRESPONDENCE ADDRESS:
;        ADDRESSEE:  No. 5646025o No. 5646025disk of No. 5646025th America, Inc.
;        STREET:  405 Lexington Avenue, 64th Floor
;        CITY:  New York
;        STATE:  New York
;        COUNTRY:  USA
;        ZIP:  10174-6401
;      COMPUTER READABLE FORM:
;        MEDIUM TYPE:  Floppy disk
;        COMPUTER:  IBM PC compatible
;        OPERATING SYSTEM:  PC-DOS/MS-DOS
;        SOFTWARE:  PatentIn Release #1.0, Version #1.30
;      CURRENT APPLICATION DATA:
;        APPLICATION NUMBER:  US/08/435,925C
;        FILING DATE:  05-MAY-1995
;        CLASSIFICATION:  435

;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lanbiris, Elias J.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  4429.000-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2794 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  283..413
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  618..696
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  718..793
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  956..1108
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  1166..1218
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  1789..1842
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  2671..2764
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1..282, 414..617, 697..717, 794..955, 1109
;      LOCATION:  ..1165, 1219..1788, 1843..2670, 2765..2791)
US-08-435-925C-1

  Query Match             74.0%;  Score 14.8;  DB 1;  Length 2794;
  Best Local Similarity   88.9%;  Pred. No. 61;
  Matches  16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAAC 18
           |||| ||||||||||||
Db    1537 GATCGCAATCCGACCAAC 1554


RESULT 7
US-08-986-485-1
; Sequence 1, Application US/08986485
; Patent No. 6046030
;    GENERAL INFORMATION:
;      APPLICANT:  WU, SHUJIAN
;      APPLICANT:  SWEET, RAYMOND
;      APPLICANT:  TRUNEH, ALEMSEGED
;      TITLE OF INVENTION:  A HUMAN LIG-1 HOMOLOG (HLIG-1)
;      NUMBER OF SEQUENCES:  8
;      CORRESPONDENCE ADDRESS:
;        ADDRESSEE:  RATNER & PRESTIA
;        STREET:  P.O. BOX 980
;        CITY:  VALLEY FORGE
;        STATE:  PA
;        COUNTRY:  USA
;        ZIP:  19482
;      COMPUTER READABLE FORM:
;        MEDIUM TYPE:  Diskette
;        COMPUTER:  IBM Compatible
;        OPERATING SYSTEM:  DOS
;        SOFTWARE:  PastSEQ for Windows Version 2.0
;      CURRENT APPLICATION DATA:
;        APPLICATION NUMBER:  US/08/986,485
```

```
;       FILING DATE:  08-DEC-1997
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  60/059,448
;       FILING DATE:  22-SEP-1997
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  PRESTIA, PAUL F
;       REGISTRATION NUMBER:  23,031
;       REFERENCE/DOCKET NUMBER:  GH-70264
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  610-407-0700
;       TELEFAX:  610-407-0701
;       TELEX:  846169
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4843 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-986-485-1

  Query Match             74.0%;  Score 14.8;  DB 3;  Length 4843;
  Best Local Similarity   88.9%;  Pred. No. 66;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCGACCAAC 18
           |||||||||| || ||||
Db     478 GATCTGAATCTGAACAAC 495


RESULT 8
US-09-385-982-497
; Sequence 497, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
;  APPLICANT: ENDEGE, WILSON O., ET AL.
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;  TITLE OF INVENTION: PRODUCTS:  II
;  FILE REFERENCE: CCDNA-260XX
;  CURRENT APPLICATION NUMBER: US/09/385,982
;  CURRENT FILING DATE: 1999-08-30
;  EARLIER APPLICATION NUMBER: 09/328,111
;  EARLIER FILING DATE: 1999-06-08
;  EARLIER APPLICATION NUMBER: 60/117,393
;  EARLIER FILING DATE: 1999-01-27
;  EARLIER APPLICATION NUMBER: 60/098,639
;  EARLIER FILING DATE: 1998-08-31
;  NUMBER OF SEQ ID NOS: 544
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 497
;    LENGTH: 587
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (1)...(587)
;    OTHER INFORMATION: n = A,T,C or G
US-09-385-982-497

  Query Match             71.0%;  Score 14.2;  DB 4;  Length 587;
  Best Local Similarity   84.2%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACC 19
           |||| || |||||||||
Db     505 GATCCGACTCGGACCAACC 523


RESULT 9
US-09-098-219B-1
; Sequence 1, Application US/09098219B
; Patent No. 6441277

; GENERAL INFORMATION:
;   APPLICANT:  Barry, Gerard
;   APPLICANT:  Cheikh, No. 6441277dine
;   APPLICANT:  Kishore, Ganesh
;   TITLE OF INVENTION:  Expression of Fructose 1,6 Bisphosphate
;   TITLE OF INVENTION:  Aldolase in Transgenic Plants
;   NUMBER OF SEQUENCES:  6
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Arnold White & Durkee
;     STREET:  P.O. Box 4433
;     CITY:  Houston
;     STATE:  TX
;     COUNTRY:  US
;     ZIP:  77210-4433
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/09/098,219B
;     FILING DATE:
;     CLASSIFICATION:  800
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 60/049,995
;     FILING DATE:  17-JUN-1997
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Kammerer, Patricia A.
;     REGISTRATION NUMBER:  29,775
;     REFERENCE/DOCKET NUMBER:  MOBT:086
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  713-787-1400
;     TELEFAX:  713-787-1440
; INFORMATION FOR SEQ ID NO:  1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  1080 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
US-09-098-219B-1

  Query Match             71.0%;  Score 14.2;  DB 4;  Length 1080;
  Best Local Similarity   84.2%;  Pred. No. 1.1e+02;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACC 19
           | ||||| |||||||| ||
Db     706 GTTCTGACTCCGACCATCC 724


RESULT 10
US-08-976-259-22
; Sequence 22, Application US/08976259
; Patent No. 6316609
;  GENERAL INFORMATION:
;    APPLICANT:  Dillon, Patrick J.
;    APPLICANT:  Choi, Gil H.
;    APPLICANT:  Welch, Rodney A.
;    TITLE OF INVENTION:  Nucleotide Sequence of Escherichia coli
; Patent No. 6316609
;    NUMBER OF SEQUENCES:  142
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;      STREET:  1100 New York Ave, N.W., Suite 600
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
```

```
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/976,259
;       FILING DATE:  Herewith
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 60/031,626 AND US 60/061,953
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Steffe, Eric K.
;       REGISTRATION NUMBER:  36,688
;       REFERENCE/DOCKET NUMBER:  1488.0740002/EKS/CBM
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 371-2600
;       TELEFAX:  (202) 371-2540
;   INFORMATION FOR SEQ ID NO:  22:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2916 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-08-976-259-22

  Query Match           71.0%;  Score 14.2;  DB 4;  Length 2916;
  Best Local Similarity   84.2%;  Pred. No. 1.3e+02;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ATCTGAATCCGACCAACCG 20
           ||||||  |||||||||  |
Db     189 ATCTGACGCCGACCAACTG 207


RESULT 11
US-08-176-620A-5
; Sequence 5, Application US/08176620A
; Patent No. 5595904
;  GENERAL INFORMATION:
;    APPLICANT:  Boulton, Teri G.
;    APPLICANT:  Cobb, Melanie H.
;    APPLICANT:  Yancopoulos, George D.
;    APPLICANT:  Nye, Steven
;    APPLICANT:  Panayotatos, Nikos
;    TITLE OF INVENTION:  A Family of Map2 Protein Kinases
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/176,620A
;      FILING DATE:  03-JAN-1994
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  6526-123
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3671 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  303..2018
US-08-176-620A-5

  Query Match           71.0%;  Score 14.2;  DB 1;  Length 3671;
  Best Local Similarity   84.2%;  Pred. No. 1.3e+02;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACC 19
           ||||| || ||| ||||||
Db     843 GATCTCAAGCCGGCCAACC 861


RESULT 12
US-08-463-862-5
; Sequence 5, Application US/08463862
; Patent No. 5776751
;  GENERAL INFORMATION:
;    APPLICANT:  Boulton, Teri G. et al.
;    TITLE OF INVENTION:  A FAMILY OF MAP2 PROTEIN KINASES
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,862
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/701,544
;      FILING DATE:  16-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  6526-049
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3671 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  303..2018
US-08-463-862-5

  Query Match           71.0%;  Score 14.2;  DB 1;  Length 3671;
  Best Local Similarity   84.2%;  Pred. No. 1.3e+02;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACC 19
           ||||| || ||| ||||||
Db     843 GATCTCAAGCCGGCCAACC 861


RESULT 13
US-08-461-985-5
; Sequence 5, Application US/08461985
```

```
; Patent No. 5872006
;  GENERAL INFORMATION:
;    APPLICANT:  Boulton, Teri G.
;    APPLICANT:  Cobb, Melanie H.
;    APPLICANT:  Yancopoulos, George D.
;    APPLICANT:  Nye, Steven
;    APPLICANT:  Panayotatos, Nikos
;    TITLE OF INVENTION:  A Family of Map2 Protein Kinases
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/461,985
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/176,620
;      FILING DATE:  03-JAN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  6526-123
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3671 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  303..2018
US-08-461-985-5

  Query Match           71.0%;  Score 14.2;  DB 2;  Length 3671;
  Best Local Similarity   84.2%;  Pred. No. 1.3e+02;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACC 19
           ||||| || ||| ||||||
Db     843 GATCTCAAGCCGGCCAACC 861


RESULT 14
US-08-458-887-5
; Sequence 5, Application US/08458887
; Patent No. 5914261
;  GENERAL INFORMATION:
;    APPLICANT:  Boulton, Teri G. et al.
;    TITLE OF INVENTION:  A FAMILY OF MAP2 PROTEIN KINASES
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,887
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/701,544
;      FILING DATE:  16-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  6526-049
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3671 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  303..2018
US-08-458-887-5

  Query Match           71.0%;  Score 14.2;  DB 2;  Length 3671;
  Best Local Similarity   84.2%;  Pred. No. 1.3e+02;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACC 19
           ||||| || ||| ||||||
Db     843 GATCTCAAGCCGGCCAACC 861


RESULT 15
US-08-390-888A-2/c
; Sequence 2, Application US/08390888A
; Patent No. 5916754
;  GENERAL INFORMATION:
;    APPLICANT:  Nichol, Stuart T.
;    APPLICANT:  Morzunov, Sergey
;    APPLICANT:  Ksiazek, Thomas G.
;    APPLICANT:  Rollin, Pierre E.
;    APPLICANT:  Spiropoulou, Christina F.
;    TITLE OF INVENTION:  THE BAYOU HANTAVIRUS AND RELATED METHODS
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  127 Peachtree Street, N.E., Suite 1200
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC Compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/390,888A
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spratt, Gwendolyn D.
;      REGISTRATION NUMBER:  36,016
;      REFERENCE/DOCKET NUMBER:  1414.623
;    TELECOMMUNICATION INFORMATION:
```

; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3677 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-390-888A-2

Query Match 71.0%; Score 14.2; DB 2; Length 3677;
Best Local Similarity 84.2%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GATCTGAATCCGACCAACC 19
|| ||| ||| ||||||||
Db 2053 GAGCTGTATCAGACCAACC 2035

Search completed: May 12, 2003, 02:38:31
Job time : 11.7857 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 02:35:34 ; Search time 18.7143 Seconds
(without alignments)
1328.082 Million cell updates/sec

Title: US-09-980-277-2
Perfect score: 20
Sequence: 1 gatctgaatccgaccaaccg 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	15.8	79.0	530	10	US-09-864-761-13599	Sequence 13599, A
c	2	15.8	79.0	1005	10	US-09-815-242-9151	Sequence 9151, Ap
	3	15.2	76.0	457	9	US-09-918-995-2350	Sequence 2350, Ap
	4	15.2	76.0	2460	9	US-09-895-298-43	Sequence 43, Appl
c	5	15.2	76.0	7914	12	US-10-095-718-3	Sequence 3, Appli
c	6	15.2	76.0	7944	12	US-10-095-718-1	Sequence 1, Appli
	7	15.2	76.0	19929	9	US-09-764-891-9967	Sequence 9967, Ap
	8	15.2	76.0	20907	9	US-09-764-891-9966	Sequence 9966, Ap
	9	15.2	76.0	536165	9	US-09-939-964-1	Sequence 1, Appli
	10	14.8	74.0	437	10	US-09-867-701-3071	Sequence 3071, Ap
	11	14.8	74.0	479	9	US-09-918-995-794	Sequence 794, App
c	12	14.8	74.0	560	10	US-09-864-761-32024	Sequence 32024, A
c	13	14.8	74.0	732	9	US-09-938-842A-1568	Sequence 1568, Ap
c	14	14.8	74.0	1985	10	US-09-864-761-15510	Sequence 15510, A
	15	14.4	72.0	459	10	US-09-770-444-455	Sequence 455, App
	16	14.4	72.0	773	10	US-09-974-300-2462	Sequence 2462, Ap
c	17	14.2	71.0	241	10	US-09-878-574-6740	Sequence 6740, Ap
	18	14.2	71.0	421	9	US-09-918-995-8698	Sequence 8698, Ap
	19	14.2	71.0	454	9	US-09-918-995-3916	Sequence 3916, Ap
c	20	14.2	71.0	466	10	US-09-864-761-14634	Sequence 14634, A
	21	14.2	71.0	485	10	US-09-917-800A-585	Sequence 585, App
	22	14.2	71.0	572	10	US-09-974-300-1878	Sequence 1878, Ap
c	23	14.2	71.0	615	10	US-09-974-300-4348	Sequence 4348, Ap
c	24	14.2	71.0	854	10	US-09-925-300-382	Sequence 382, App
	25	14.2	71.0	1006	10	US-09-770-445-227	Sequence 227, App
	26	14.2	71.0	1080	10	US-09-923-109-1	Sequence 1, Appli
c	27	14.2	71.0	1275	10	US-09-974-300-917	Sequence 917, App
	28	14.2	71.0	1372	9	US-09-938-842A-2894	Sequence 2894, Ap
c	29	14.2	71.0	1407	10	US-09-898-932-1	Sequence 1, Appli
	30	14.2	71.0	2179	10	US-09-867-550-1891	Sequence 1891, Ap
	31	14.2	71.0	2318	9	US-10-260-046-29	Sequence 29, Appl
	32	14.2	71.0	2604	9	US-09-938-842A-458	Sequence 458, App
	33	14.2	71.0	2632	12	US-10-044-090-312	Sequence 312, App
	34	14.2	71.0	2916	10	US-09-956-004-22	Sequence 22, Appl
	35	14.2	71.0	5510	12	US-10-044-090-353	Sequence 353, App
	36	14.2	71.0	10846	10	US-09-923-109-5	Sequence 5, Appli
	37	14.2	71.0	10900	10	US-09-923-109-6	Sequence 6, Appli
	38	14.2	71.0	21129	9	US-10-091-504-1734	Sequence 1734, Ap
	39	14.2	71.0	21129	10	US-09-764-869-1734	Sequence 1734, Ap
	40	14.2	71.0	147309	10	US-09-742-312-3	Sequence 3, Appli
	41	14.2	71.0	684973	10	US-09-263-959-1	Sequence 1, Appli
	42	13.8	69.0	657	10	US-09-891-171-4	Sequence 4, Appli
	43	13.8	69.0	197	10	US-09-983-965-4349	Sequence 4349, Ap
c	44	13.8	69.0	231	10	US-09-960-352-14201	Sequence 14201, A
	45	13.8	69.0	271	10	US-09-923-876-4684	Sequence 4684, Ap

ALIGNMENTS

RESULT 1
US-09-864-761-13599
; Sequence 13599, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecaica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30

```
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 13599
;    LENGTH: 530
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    OTHER INFORMATION: MAP TO AC023314.3
;    OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
US-09-864-761-13599

  Query Match             79.0%;  Score 15.8;  DB 10;  Length 530;
  Best Local Similarity   89.5%;  Pred. No. 1.1e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 ATCTGAATCCGACCAACCG 20
          |||||||||||||||| |
Db      2 ATGTGAATCCGACCAACTG 20


RESULT 2
US-09-815-242-9151/c
; Sequence 9151, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9151
;    LENGTH: 1005
;    TYPE: DNA
;    ORGANISM: Streptococcus pneumoniae
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1)...(1005)
US-09-815-242-9151

  Query Match             79.0%;  Score 15.8;  DB 10;  Length 1005;
  Best Local Similarity   89.5%;  Pred. No. 1.1e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 ATCTGAATCCGACCAACCG 20
          |||||||||||||||| |||
Db    596 ATCTGAATCCGACGAACTG 578


RESULT 3
US-09-918-995-2350
; Sequence 2350, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2350
;    LENGTH: 457
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (1)...(457)
;    OTHER INFORMATION: n = A,T,C or G
US-09-918-995-2350

  Query Match             76.0%;  Score 15.2;  DB 9;  Length 457;
  Best Local Similarity   85.0%;  Pred. No. 2.2e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GATCTGAATCCGACCAACCG 20
          || ||||||| |||||||| |
Db    142 GAACTGAATCAGACCAACTG 161


RESULT 4
US-09-895-298-43
; Sequence 43, Application US/09895298
; Publication No. US20030078405A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 47 Human Secreted Proteins
;  FILE REFERENCE: PZ035P1
;  CURRENT APPLICATION NUMBER: US/09/895,298
;  CURRENT FILING DATE:  2001-07-02
;  PRIOR APPLICATION NUMBER: 09/591,16
;  PRIOR FILING DATE: 2000-06-09
;  PRIOR APPLICATION NUMBER: PCT/US99/29950
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: 60/113,006
;  PRIOR FILING DATE: 1998-12-18
;  PRIOR APPLICATION NUMBER: 60/112,809
;  PRIOR FILING DATE: 1998-12-17
;  NUMBER OF SEQ ID NOS: 231
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
;    LENGTH: 2460
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-09-895-298-43

  Query Match             76.0%;  Score 15.2;  DB 9;  Length 2460;
  Best Local Similarity   85.0%;  Pred. No. 2.2e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GATCTGAATCCGACCAACCG 20
          || ||||||| |||||||| |
```

```
Db    2175 GAACTGAATCAGACCAACTG 2194


RESULT 5
US-10-095-718-3/c
; Sequence 3, Application US/10095718
; Patent No. US20020131956A1
; GENERAL INFORMATION:
;  APPLICANT: Walsh, Christopher
;  APPLICANT:  Chao, Hengjun
;  APPLICANT:  Burstein, Haim
;  APPLICANT:  Lynch, Carmel
;  APPLICANT:  Stepan, Tony
;  APPLICANT:  Munson, Keith
;  TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and
;  TITLE OF INVENTION:  Methods of Using the Same
;  FILE REFERENCE: 35052/204375
;  CURRENT APPLICATION NUMBER: US/10/095,718
;  CURRENT FILING DATE:  2002-03-12
;  PRIOR APPLICATION NUMBER: 09/689,430
;  PRIOR FILING DATE: 2001-08-22
;  PRIOR APPLICATION NUMBER: 60/158,780
;  PRIOR FILING DATE: 1999-10-12
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PastSEQ for Windows Version 4.0
; SEQ ID NO 3
;    LENGTH: 7914
;    TYPE: DNA
;    ORGANISM: Artificial Sequence
;    FEATURE:
;    OTHER INFORMATION: rAAV vector with canine B-domain deleted factor
;    OTHER INFORMATION: VIII
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (435)...(4730)
US-10-095-718-3

  Query Match           76.0%;  Score 15.2;  DB 12;  Length 7914;
  Best Local Similarity   85.0%;  Pred. No. 2.2e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCGACCAACCG 20
           |||||| |||||| ||| ||
Db    4890 GATCTGGATCCGACAAACG 4871


RESULT 6
US-10-095-718-1/c
; Sequence 1, Application US/10095718
; Patent No. US20020131956A1
; GENERAL INFORMATION:
;  APPLICANT: Walsh, Christopher
;  APPLICANT:  Chao, Hengjun
;  APPLICANT:  Burstein, Haim
;  APPLICANT:  Lynch, Carmel
;  APPLICANT:  Stepan, Tony
;  APPLICANT:  Munson, Keith
;  TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and
;  TITLE OF INVENTION:  Methods of Using the Same
;  FILE REFERENCE: 35052/204375
;  CURRENT APPLICATION NUMBER: US/10/095,718
;  CURRENT FILING DATE:  2002-03-12
;  PRIOR APPLICATION NUMBER: 09/689,430
;  PRIOR FILING DATE: 2001-08-22
;  PRIOR APPLICATION NUMBER: 60/158,780
;  PRIOR FILING DATE: 1999-10-12
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PastSEQ for Windows Version 4.0
; SEQ ID NO 1
;    LENGTH: 7944
;    TYPE: DNA
;    ORGANISM: Artificial Sequence
;    FEATURE:
;    OTHER INFORMATION: Plasmid pDL66 encoding Homo sapiens BDD FVIII
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (420)...(4835)
US-10-095-718-1

  Query Match           76.0%;  Score 15.2;  DB 12;  Length 7944;
  Best Local Similarity   85.0%;  Pred. No. 2.2e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCGACCAACCG 20
           |||||| |||||| ||| ||
Db    4920 GATCTGGATCCGACAAACG 4901


RESULT 7
US-09-764-891-9967
; Sequence 9967, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC006
;  CURRENT APPLICATION NUMBER: US/09/764,891
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 10231
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9967
;    LENGTH: 19929
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-09-764-891-9967

  Query Match           76.0%;  Score 15.2;  DB 9;  Length 19929;
  Best Local Similarity   85.0%;  Pred. No. 2.3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCGACCAACCG 20
           |||||| |||||| ||| ||
Db   18965 GAACTGAATCAGACCAACTG 18984


RESULT 8
US-09-764-891-9966
; Sequence 9966, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC006
;  CURRENT APPLICATION NUMBER: US/09/764,891
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 10231
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9966
;    LENGTH: 20907
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: SITE
;    LOCATION: (4477)
;    OTHER INFORMATION: n equals a,t,g, or c
;    NAME/KEY: SITE
;    LOCATION: (4480)
;    OTHER INFORMATION: n equals a,t,g, or c
;    NAME/KEY: SITE
;    LOCATION: (4483)
;    OTHER INFORMATION: n equals a,t,g, or c
US-09-764-891-9966

  Query Match           76.0%;  Score 15.2;  DB 9;  Length 20907;
```

```
  Best Local Similarity   85.0%;  Pred. No. 2.3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACCG 20
           || ||||||| ||||||| |
Db   18962 GAACTGAATCAGACCAACTG 18981


RESULT 9
US-09-939-964-1
; Sequence 1, Application US/09939964
; Publication No. US20030054522A1
; GENERAL INFORMATION:
;  APPLICANT: Rosenthal, Andre
;  APPLICANT:  Freiberg, Christoph
;  APPLICANT:  Perret, Xavier Philippe
;  APPLICANT:  Broughton, William John
;  TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic
;  TITLE OF INVENTION:  Plasmid
;  FILE REFERENCE: CARP0068
;  CURRENT APPLICATION NUMBER: US/09/939,964
;  CURRENT FILING DATE:  2001-08-27
;  PRIOR APPLICATION NUMBER: 09/214,808
;  PRIOR FILING DATE: 1999-06-22
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 536165
;   TYPE: DNA
;   ORGANISM: Rhizobium
US-09-939-964-1

  Query Match             76.0%;  Score 15.2;  DB 9;  Length 536165;
  Best Local Similarity   85.0%;  Pred. No. 2.3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACCG 20
           ||| ||||||||||| ||||
Db  222412 GATGAGAATCCGACCGACCG 222431


RESULT 10
US-09-867-701-3071
; Sequence 3071, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
;  APPLICANT: Aglate, Paul A.
;  APPLICANT:  Jones, Robert
;  APPLICANT:  Harlocker, Susan L.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE: 210121.497
;  CURRENT APPLICATION NUMBER: US/09/867,701
;  CURRENT FILING DATE:  2001-05-29
;  NUMBER OF SEQ ID NOS: 10912
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3071
;   LENGTH: 437
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-867-701-3071

  Query Match             74.0%;  Score 14.8;  DB 10;  Length 437;
  Best Local Similarity   88.9%;  Pred. No. 3.4e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAAC 18
           |||||||||| || ||||
Db     309 GATCTGAATCTGAACAAC 326


RESULT 11
US-09-918-995-794
; Sequence 794, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 794
;   LENGTH: 479
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(479)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-794

  Query Match             74.0%;  Score 14.8;  DB 9;  Length 479;
  Best Local Similarity   88.9%;  Pred. No. 3.4e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 ATCTGAATCCGACCAACC 19
           | ||||||||||| |||||
Db     434 ACCTGAATCCGATCAACC 451


RESULT 12
US-09-864-761-32024/c
; Sequence 32024, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL I
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomics-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
```

```
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32024
;    LENGTH: 560
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    OTHER INFORMATION: MAP TO AF127577.1
;    OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.98
;    OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  1
;    OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.99
;    OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.99
;    OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
;    OTHER INFORMATION: NT HIT: AL163207.2, EVALUE 0.00e+00
;    OTHER INFORMATION: EST_HUMAN HIT: BE082541.1, EVALUE 0.00e+00
;    OTHER INFORMATION: SWISSPROT HIT: P48552, EVALUE 1.00e-106
US-09-864-761-32024

  Query Match             74.0%;  Score 14.8;  DB 10;  Length 560;
  Best Local Similarity   88.9%;  Pred. No. 3.4e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 ATCTGAATCCGACCAACC 19
           ||| ||||| ||||||||
Db     229 ATCAGAATCTGACCAACC 212


RESULT 13
US-09-938-842A-1568/c
; Sequence 1568, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
;  APPLICANT: Harper, Jeff
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Wang, Xun
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
;  TITLE OF INVENTION:  SAME, AND METHODS OF USE
;  FILE REFERENCE: SCRIP1300-3
;  CURRENT APPLICATION NUMBER: US/09/938,842A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 60/227,866
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: US 60/264,647
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/300,111
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1568
;    LENGTH: 732
;    TYPE: DNA
;    ORGANISM: Arabidopsis thaliana
US-09-938-842A-1568

  Query Match             74.0%;  Score 14.8;  DB 9;  Length 732;
  Best Local Similarity   88.9%;  Pred. No. 3.4e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 ATCTGAATCCGACCAACC 19
           |||||| ||||||| |||
Db     504 ATCTGATTCCGACCCACC 487


RESULT 14
US-09-864-761-15510/c
; Sequence 15510, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 15510
;    LENGTH: 1985
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    OTHER INFORMATION: MAP TO AF127577.1
;    OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.98
;    OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  1
;    OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.99
;    OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.99
;    OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
US-09-864-761-15510

  Query Match             74.0%;  Score 14.8;  DB 10;  Length 1985;
  Best Local Similarity   88.9%;  Pred. No. 3.5e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 ATCTGAATCCGACCAACC 19
           ||| ||||| ||||||||
Db     600 ATCAGAATCTGACCAACC 583
```

```
RESULT 15
US-09-770-444-455
; Sequence 455, Application US/09770444
; Patent No. US20020023280A1
; GENERAL INFORMATION:
;  APPLICANT: Gorlach, Jorn
;  APPLICANT:  An, Yong-Qiang
;  APPLICANT:  Hamilton, Carol M.
;  APPLICANT:  Price, Jennifer L.
;  APPLICANT:  Raines, Tracy M.
;  APPLICANT:  Yu, Yang
;  APPLICANT:  Kameaka, Joshua G.
;  APPLICANT:  Page, Amy
;  APPLICANT:  Matthew, Abraham V.
;  APPLICANT:  Ledford, Brooke L.
;  APPLICANT:  Woessner, Jeffrey P.
;  APPLICANT:  Haas, William David
;  APPLICANT:  Garcia, Carlos A.
;  APPLICANT:  Kricker, Maja
;  APPLICANT:  Slader, Ted
;  APPLICANT:  Davis, Keith R.
;  APPLICANT:  Allen, Keith
;  APPLICANT:  Hoffman, Neil
;  APPLICANT:  Hurban, Patrick
;  TITLE OF INVENTION: Expressed Sequences of Arabidopsis
;  TITLE OF INVENTION:  thaliana
;  FILE REFERENCE: 2027 (PARA-016PRV)
;  CURRENT APPLICATION NUMBER: US/09/770,444
;  CURRENT FILING DATE:  2001-01-26
;  PRIOR APPLICATION NUMBER: 60/178,502
;  PRIOR FILING DATE: 2000-01-27
;  NUMBER OF SEQ ID NOS: 999
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 455
;   LENGTH: 459
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-770-444-455

  Query Match             72.0%;  Score 14.4;  DB 10;  Length 459;
  Best Local Similarity   93.8%;  Pred. No. 5.4e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 ATCTGAATCCGACCAA 17
           ||||||| ||||||||
Db      56 ATCTGAAGCCGACCAA 71


Search Completed: May 12, 2003, 04:49:50
Job time : 130.714 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 01:04:28 ; Search time 244.143 Seconds
(without alignments)
1326.722 Million cell updates/sec

Title: US-09-980-277-2
Perfect score: 20
Sequence: 1 gatctgatccgaccaaccg 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	18.4	92.0	687	17	BH210086	BH210086 Sm1-56J20
2	16.8	84.0	838	17	BH526706	BH526706 BOGNZ32TR
3	16.4	82.0	528	17	TA11G10P	AL451692 T. brucei
4	16.4	82.0	640	17	AQ638867	AQ638867 927P1-8A1
5	16.4	82.0	696	12	BF621944	BF621944 HVSMEa000
6	16	80.0	647	17	BH516307	BH516307 BOGKX42TF
	7	16	80.0	745	17	BH077414	BH077414 RPCI-24-3
	8	16	80.0	909	17	BH130484	BH130484 RMTOL69TF
c	9	15.8	79.0	184	13	BM024534	BM024534 fu48g01.y
	10	15.8	79.0	234	10	BB398677	BB398677 BB398677
	11	15.8	79.0	256	10	BB357524	BB357524 BB357524
	12	15.8	79.0	285	14	BQ236459	BQ236459 TaK05031E
	13	15.8	79.0	300	17	BH224590	BH224590 1006119H1
c	14	15.8	79.0	302	13	BM443097	BM443097 EBro02_SQ
	15	15.8	79.0	328	13	BB297513	BB297513 BB297513
c	16	15.8	79.0	360	9	AJ462699	AJ462699 AJ462699
	17	15.8	79.0	362	17	AZ921320	AZ921320 1006029P0
	18	15.8	79.0	390	9	AA544509	AA544509 vr66611.r
	19	15.8	79.0	406	17	BH237038	BH237038 ATZUB88TR
	20	15.8	79.0	436	17	BH224579	BH224579 1006119H0
	21	15.8	79.0	454	17	BH225175	BH225175 1006124X0
c	22	15.8	79.0	458	10	BB745434	BB745434 BB745434
	23	15.8	79.0	533	10	BE446503	BE446503 WHE1456_E
c	24	15.8	79.0	533	12	BG739245	BG739245 fp68f08.y
c	25	15.8	79.0	554	13	BI981975	BI981975 fu53f11.y
c	26	15.8	79.0	557	13	BM103560	BM103560 fv19h09.y
	27	15.8	79.0	560	13	BJ221092	BJ221092 BJ221092
c	28	15.8	79.0	570	13	BI350830	BI350830 fz39g01.y
c	29	15.8	79.0	596	13	BI706725	BI706725 fq09c12.y
c	30	15.8	79.0	600	17	BH236984	BH236984 ATZUB87TR
	31	15.8	79.0	606	10	AV835595	AV835595 AV835595
	32	15.8	79.0	620	13	BJ466527	BJ466527 BJ466527
c	33	15.8	79.0	621	10	BE189862	BE189862 db61f03.y
c	34	15.8	79.0	626	17	BH743324	BH743324 gz72d06.g
c	35	15.8	79.0	628	9	AL797621	AL797621 AL797621
c	36	15.8	79.0	631	9	AL680077	AL680077 AL680077
	37	15.8	79.0	637	17	BH449707	BH449707 BOGNT77TR
c	38	15.8	79.0	643	12	BF937817	BF937817 fm68d03.y
c	39	15.8	79.0	643	17	BH236968	BH236968 ATZUC25TR
c	40	15.8	79.0	653	10	AW422560	AW422560 fj94h04.y
c	41	15.8	79.0	655	9	AL655921	AL655921 AL655921
c	42	15.8	79.0	659	13	BJ071265	BJ071265 BJ071265
c	43	15.8	79.0	662	13	BJ064162	BJ064162 BJ064162
	44	15.8	79.0	663	17	BH589281	BH589281 BOGMF78TF
c	45	15.8	79.0	667	13	BJ097224	BJ097224 BJ097224

ALIGNMENTS

RESULT 1
BH210086
LOCUS BH210086 687 bp DNA linear GSS 24-OCT-2001
DEFINITION Sm1-56J20.TF Sm1 Schistosoma mansoni genomic clone sm1-56J20, DNA sequence.
ACCESSION BH210086
VERSION BH210086.1 GI:16389299
KEYWORDS GSS.
SOURCE Schistosoma mansoni.
ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE 1 (bases 1 to 687)
AUTHORS Shetty,J., Simpson,A., Malek,J., Koo,H., LoVerde,P.T. and El-Sayed ,N.M.
TITLE Use of end sequences from Schistosoma mansoni (Puerto Rico strain) Sm1 BAC library for gene discovery and map construction
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: Sm1-56J20.TR
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
lo.edu).
Seq primer: M13 For
Class: BAC ends.

```
FEATURES             Location/Qualifiers
     source          1. .687
                     /organism="Schistosoma mansoni"
                     /strain="Puerto Rico"
                     /db_xref="taxon:6183"
                     /clone="Sm1-56J20"
                     /clone_lib="Sm1"
                     /note="Vector: pBeloBAC11; Site_1: Hin dIII; Constructed
                     in the laboratory of Dr. Denis Le Paslier at the Fondation
                     Jean Dausset, CEPH, Paris, France. Briefly, Schistosoma
                     mansoni agarose embedded DNA was partially digested with
                     Hin dIII. High molecular weight fragments were ligated in
                     pBeloBAC11 digested with Hin dIII. The average insert size
                     is 100 kb. Total clone coverage: approx. 7.95 X the
                     haploid genome. Further information can be found in Le
                     Paslier et al. (2000) Construction and characterization of
                     a Schistosoma mansoni bacterial artificial chromosome
                     library. Genomics 65: 87-94."
BASE COUNT       218 a    101 c    152 g    216 t
ORIGIN

  Query Match          92.0%;  Score 18.4;  DB 17;  Length 687;
  Best Local Similarity   95.0%;  Pred. No. 48;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACCG 20
           |||||||||||||||| |||
Db     348 GATCTGAATCCGACCATCCG 367


RESULT 2
BH526706
LOCUS       BH526706                838 bp    DNA     linear   GSS 13-DEC-2001
DEFINITION  BOGNI32TR BoGN Brassica oleracea genomic clone BOGNI32, DNA
            sequence.
ACCESSION   BH526706
VERSION     BH526706.1  GI:17734791
KEYWORDS    GSS.
SOURCE      Brassica oleracea.
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 838)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOGNI32TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .838
                     /organism="Brassica oleracea"
                     /strain="PO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGNI32"
                     /clone_lib="BOGN"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       277 a    166 c    112 g    283 t
ORIGIN

  Query Match          84.0%;  Score 16.8;  DB 17;  Length 838;
  Best Local Similarity   90.0%;  Pred. No. 3.5e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACCG 20
           |||||||||||| || |||
Db     543 GATCTGAATCCGAACATCCG 562


RESULT 3
TAllG10P
LOCUS       TAllG10P                528 bp    DNA     linear   GSS 13-DEC-2000
DEFINITION  T. brucei sheared genomic DNA clone 11g10, forward sequence,
            genomic survey sequence.
ACCESSION   AL451692
VERSION     AL451692.1  GI:11831715
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
            Chillingsworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
            Melville,S.E., Rajandream,M.A. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
            project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
            Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
            nh1@sanger.ac.uk
COMMENT     Constructed at the Institute for Genomic Research (TIGR),
            Rockville, MD. Genomic DNA isolated from a cloned population of
            Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
            to give a tight size distribution (
            4 kb). The Y + i method used for the library construction is
            described in detail in Smith, H. and Venter, J.C. (Making small
            insert libraries for whole genome shotgun sequencing projects.  In
            Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999).
            Email: nelsayed@tigr.org
            Details of T. brucei sequencing at the Sanger Centre are available
            at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Trypanosoma brucei"
                     /strain="TREU927"
                     /db_xref="taxon:5691"
                     /clone="11g10"
BASE COUNT       122 a    155 c    118 g    133 t
ORIGIN

  Query Match          82.0%;  Score 16.4;  DB 17;  Length 528;
  Best Local Similarity   94.4%;  Pred. No. 4.8e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 TCTGAATCCGACCAACCG 20
           ||||| ||||||||||||
Db     208 TCTGAACCCGACCAACCG 225


RESULT 4
AQ638867
LOCUS       AQ638867                640 bp    DNA     linear   GSS 08-JUL-1999
DEFINITION  927F1-8A1.TV 927F1 Trypanosoma brucei genomic clone 927F1-8A1, DNA
            sequence.
ACCESSION   AQ638867
VERSION     AQ638867.1  GI:5115577
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 640)
  AUTHORS   El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
            Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.
            , Fraser,C. and Adams,M.
  TITLE     Determination of clone end sequences from Trypanosoma brucei TREU
```

```
                927/4 P1 library
  JOURNAL       Unpublished (1999)
COMMENT         Other_GSSs: 927P1-BA1.TP
                Contact: Najib M. El-Sayed
                Department of Eukaryotic Genomics
                The Institute for Genomic Research
                9712 Medical Center Dr., Rockville, MD 20850, USA
                Tel: 301 838 0200
                Fax: 301 838 0208
                Email: nelsayed@tigr.org
                For clone/filter availability, please contact Sara Melville
                (sm160@mole.bio.cam.ac.uk). P1 end sequences search page:
                http://www.tigr.org/tdb/mdb/tbdb/.
                Seq primer: T7
                Class: P1 ends.
FEATURES                Location/Qualifiers
     source             1..640
                        /organism="Trypanosoma brucei"
                        /strain="TREU927/4"
                        /db_xref="taxon:5691"
                        /clone="927P1-BA1"
                        /clone_lib="927P1"
                        /note="Vector: pAD10SacBII; Site_1: Bam H1; Constructed by
                        Sara Melville, University of Cambridge, UK and Nancy
                        Shepherd, Dupont Merck, Wilmington, DE.  Genomic DNA was
                        isolated from Trypanosoma brucei (stock TREU927/4) and
                        partially digested with Sau 3A1. DNA fragments were cloned
                        into the Bam H1 site of pAD10SacBII vector (Genbank
                        accession U09128).  The average insert size is 65 Kb.
                        Coverage: approx 4.4 X the haploid non-minichromosomal
                        genome"
BASE COUNT      157 a    181 c    155 g    147 t
ORIGIN

  Query Match           82.0%;  Score 16.4;  DB 17;  Length 640;
  Best Local Similarity  94.4%;  Pred. No. 5.1e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 TCTGAATCCGACCAACCG 20
          |||||| ||||||||||
Db    379 TCTGAACCCGACCAACCG 396


RESULT 5
BF621944
LOCUS       BF621944                696 bp    mRNA    linear   EST 17-OCT-2001
DEFINITION  HVSMEa0001F16f Hordeum vulgare seedling shoot EST library
            HVcDNA0001 (Cold stress) Hordeum vulgare cDNA clone HVSMEa0001F16f,
            mRNA sequence.
ACCESSION   BF621944
VERSION     BF621944.2  GI:13079578
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 696)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Oates,R., Choi,D.W.
            , Fenton,R.D. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Morex cold-stressed seedling shoot cDNA
            library
  JOURNAL   Unpublished (2001)
COMMENT     On Dec 18, 2000 this sequence version replaced gi:11885678.
            Contact: Wing RA
            Clemson University Genomics Institute
            ClemSon University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases - 167
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence stop: 485.
FEATURES                Location/Qualifiers
     source             1..696
                        /organism="Hordeum vulgare"
                        /cultivar="Morex"
                        /db_xref="taxon:4513"
                        /clone="HVSMEa0001F16f"
                        /clone_lib="Hordeum vulgare seedling shoot EST library
                        HVcDNA0001 (Cold stress)"
                        /tissue_type="Seedling shoot"
                        /lab_host="TJC121"
                        /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1;
                        Seeds were surface sterilized then germinated under axenic
                        conditions in the dark at room temperature on filter paper
                        with water, nystatin and cefotaxime in covered
                        crystallization dishes. Five-day old seedlings were
                        incubated at 5oC for 2 days. Shoots were then harvested,
                        total RNA was prepared, poly(A) RNA was purified, one
                        primary unamplified cDNA library was made, and 600000 pfu
                        were in vivo excised to give pBluescript SK(-) cDNA
                        phagemids. These steps were performed in the TJ Close
                        laboratory at the University of California, Riverside
                        (Choi, Close, Fenton). Phagemids were plated and picked at
                        the Clemson University Genomics Institute (CUGI) (Begum,
                        Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations
                        , DNA sequencing and sequence analysis were performed at
                        CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main
                        ). The sequence has been trimmed to remove vector sequence
                        and contains a minimum of 100 bases of phred value 20 or
                        above. For more details on library preparation and
                        sequence analysis see
                        http://www.genome.clemson.edu/projects/barley. To order
                        this clone see http://www.genome.clemson.edu/orders Also
                        see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                        Genetically and physically anchored EST resources for
                        barley genomics. Barley Genetics Newsletter 31:29-30.
                        (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT      183 a    146 c    197 g    170 t
ORIGIN

  Query Match           82.0%;  Score 16.4;  DB 12;  Length 696;
  Best Local Similarity  94.4%;  Pred. No. 5.2e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GATCTGAATCCGACCAAC 18
          |||||||| |||||||||
Db    488 GATCTGAAACCGACCAAC 505


RESULT 6
BH516307
LOCUS       BH516307                647 bp    DNA     linear   GSS 13-DEC-2001
DEFINITION  BOGKX42TF BOGK Brassica oleracea genomic clone BOGKX42, DNA
            sequence.
ACCESSION   BH516307
VERSION     BH516307.1  GI:17724397
KEYWORDS    GSS.
SOURCE      Brassica oleracea.
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 647)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOGKX42TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
```

```
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES                Location/Qualifiers
     source             1..647
                        /organism="Brassica oleracea"
                        /strain="TO1000DH3"
                        /db_xref="taxon:3712"
                        /clone="BOGK42"
                        /clone_lib="BOGK"
                        /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                        genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      147 a    205 c    110 g    185 t
ORIGIN

  Query Match             80.0%;  Score 16;  DB 17;  Length 647;
  Best Local Similarity   100.0%;  Pred. No. 8.3e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 TCTGAATCGGACCAAC 18
           ||||||||||||||||
Db     565 TCTGAATCGGACCAAC 580


RESULT 7
BH077414
LOCUS       BH077414                745 bp    DNA     linear   GSS 18-JUL-2001
DEFINITION  RPCI-24-377I16.TV RPCI-24 Mus musculus genomic clone RPCI-24-377I16
            , DNA sequence.
ACCESSION   BH077414
VERSION     BH077414.1  GI:14897011
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 745)
  AUTHORS   Zhao,S., Nierman,W., Mmlek,J., Shatsman,S., Akiret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-377I16.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 377  row: I  column: 16
            Seq primer: T7
            Class: BAC ends.
FEATURES                Location/Qualifiers
     source             1..745
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /clone="RPCI-24-377I16"
                        /clone_lib="RPCI-24"
                        /sex="Male"
                        /cell_type="Spleen/Brain"
                        /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                        RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                        library was  cloned in the pTARBAC1 cloning vector at the
                        BamH1 sites using MboI  partially digested male C57BL/6J
                        DNA."
BASE COUNT      152 a    236 c    144 g    213 t
ORIGIN

  Query Match             80.0%;  Score 16;  DB 17;  Length 745;
  Best Local Similarity   100.0%;  Pred. No. 8.7e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 CTGAATCGGACCAACC 19
           ||||||||||||||||
Db     730 CTGAATCGGACCAACC 745


RESULT 8
BH130484
LOCUS       BH130484                909 bp    DNA     linear   GSS 07-AUG-2001
DEFINITION  ENTOL69TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   BH130484
VERSION     BH130484.1  GI:15088953
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 909)
  AUTHORS   Loftus,B., Wang,S., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            Class: shotgun
            High quality sequence start: 48
            High quality sequence stop: 770.
FEATURES                Location/Qualifiers
     source             1..909
                        /organism="Entamoeba histolytica"
                        /strain="HM1:IMSS"
                        /db_xref="taxon:5759"
                        /clone_lib="Entamoeba histolytica Sheared DNA"
                        /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                        Institute for Genomic Research (TIGR), Rockville, MD.
                        Genomic DNA isolated from broth cultures of E. histolytica
                        using a method described by Clark and Diamond (Clark,
                        C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
                        method for isolate identification.  Exp. Parasitol.
                        77:450.). The DNA was mechanically sheared to give a
                        tight size distribution (~2 kb). The v + i method used for
                        the library  construction is described in detail in Smith,
                        H.O. and Venter, J.C.  (Making small insert libraries for
                        whole genome shotgun sequencing  projects. In Genome
                        Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                        Barell, Oxford University Press, 1999)."
BASE COUNT      432 a     91 c    142 g    244 t
ORIGIN

  Query Match             80.0%;  Score 16;  DB 17;  Length 909;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ATCTGAATCGGACCAA 17
           ||||||||||||||||
Db      33 ATCTGAATCGGACCAA 48
```

```
RESULT 9
BM024534/c
LOCUS       BM024534                184 bp    mRNA    linear   EST 26-JUL-2002
DEFINITION  fu48g01.y1 zebrafish adult brain Danio rerio cDNA clone 5332800 5',
            similar to SW:GBB1_XENLA P79959 GUANINE NUCLEOTIDE-BINDING PROTEIN
            G(I)/G(S)/G(T) BETA SUBUNIT 1 ;, mRNA sequence.
ACCESSION   BM024534
VERSION     BM024534.1  GI:16538890
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 184)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,P., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Kitter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: John Ngai. cDNA Library Arrayed by:
            Matthew Clark. DNA Sequencing by: Washington University Genome
            Sequencing Center Clone distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            ResourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Trace considered overall poor quality
            Seq primer: T7
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1..184
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="5332800"
                     /clone_lib="zebrafish adult brain"
                     /sex="Mixed male and female"
                     /tissue_type="brain"
                     /dev_stage="adult"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pZIPLOX; Site_1: NotI; Site_2: SalI;
                     Original library was constructed in lambdaZIPLOX.  Mass
                     excision of the cDNA library was performed to yield
                     pZIPLOX plasmids. Insert check was done in original
                     library."
BASE COUNT       48 a     43 c     51 g     42 t
ORIGIN

  Query Match          79.0%;  Score 15.8;  DB 13;  Length 184;
  Best Local Similarity   89.5%;  Pred. No. 6.9e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 ATCTGAATCCGACCAACCG 20
           ||||||||||| ||||||||
Db      69 ATCTGAATCCGCCCAACAG 51


RESULT 10
BB398677
LOCUS       BB398677                234 bp    mRNA    linear   EST 15-JUL-2000
DEFINITION  BB398677 RIKEN full-length enriched, ES cells Mus musculus cDNA
            clone C330009I19 3', mRNA sequence.
ACCESSION   BB398677
VERSION     BB398677.1  GI:9218073
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 234)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K.K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa.Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..234
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="C330009I19"
                     /clone_lib="RIKEN full-length enriched, ES cells"
                     /cell_type="ES cells"
                     /lab_host="SOLR"
                     /note="Site_1: XhoI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 5.0 and subtraction to Rot = 25.0. Second strand
                     cDNA was prepared with the primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']."
BASE COUNT       76 a     59 c     29 g     70 t
ORIGIN

  Query Match          79.0%;  Score 15.8;  DB 10;  Length 234;
  Best Local Similarity   89.5%;  Pred. No. 7.4e+02;
```

```
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACC 19
           | ||||||| ||||||
Db      54 GCTCTGAATCCTACCAACC 72


RESULT 11
BB357524
LOCUS       BB357524                256 bp    mRNA    linear   EST 12-JUL-2000
DEFINITION  BB357524 RIKEN full-length enriched, adult male corpus striatum Mus
            musculus cDNA clone C030021K24 3' similar to X14897 Mouse fosB mRNA
            , mRNA sequence.
ACCESSION   BB357524
VERSION     BB357524.1  GI:9069352
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 256)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,R., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..256
                     /organism="Mus musculus"
                     /strain="C57SL/SJ"
                     /db_xref="taxon:10090"
                     /clone="C030021K24"
                     /clone_lib="RIKEN full-length enriched, adult male corpus
                     striatum"
                     /sex="male"
                     /tissue_type="corpus striatum"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer (5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTVN 3'), cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot - 10.0 and subtraction to Rot - 185.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
BASE COUNT       75 a    58 c    34 g    69 t
ORIGIN

  Query Match             79.0%;  Score 15.8;  DB 10;  Length 256;
  Best Local Similarity   89.5%;  Pred. No. 7.7e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACC 19
           | || | ||||||||||||
Db     119 GATTTAAATCCGACCAACC 137


RESULT 12
BQ236459
LOCUS       BQ236459                285 bp    mRNA    linear   EST 03-MAY-2002
DEFINITION  TaE05031E01F TaE05 Triticum aestivum cDNA clone TaE05031E01F, mRNA
            sequence.
ACCESSION   BQ236459
VERSION     BQ236459.1  GI:20432335
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 285)
  AUTHORS   Cloutier,S.
  TITLE     Wheat functional genomics - Glenlea developing seeds cDNA libraries
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@em.agr.ca
            was cloned directionally, not all sequences generated with reverse
            primer were from the 5' end (same with forward primer and 3' end).
            Average insert size is >2.0 kb
            Plate: 031  row: E  column: 01
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1..285
                     /organism="Triticum aestivum"
                     /cultivar="Glenlea"
                     /db_xref="taxon:4565"
                     /clone="TaE05031E01F"
                     /clone_lib="TaE05"
                     /tissue_type="developing seeds"
                     /dev_stage="5 days after anthesis"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pSPORT-P (Invitrogen Technologies); Site_1:
                     NotI; Site_2: MluI; mRNA obtained from wheat seeds of
                     cultivar Glenlea 5 days post-anthesis"
BASE COUNT       91 a    97 c    40 g    57 t
ORIGIN

  Query Match             79.0%;  Score 15.8;  DB 14;  Length 285;
```

```
  Best Local Similarity   89.5%;  Pred. No. 8e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACC 19
           |||| |||||||||| ||||
Db     204 GATCCGAATCCGACGAACC 222


RESULT 13
BH224590
LOCUS       BH224590                 300 bp    DNA     linear   GSS 08-NOV-2001
DEFINITION  1006119H11.2E1_y2 1006 - RescueMu Grid G Zea mays genomic, DNA
            sequence.
ACCESSION   BH224590
VERSION     BH224590.1  GI:16821698
KEYWORDS    GSS.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   Walbot,V.
  TITLE     Maize genomic sequences found using engineered RescueMu transposon
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Possible ligation site of ends cut by 2 different endonucleases.
            Reverse complemented post-ligation sequence from source sequence.
            Plate: 1006119  row: 44
            Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1..300
                     /organism="Zea mays"
                     /cultivar="mixed background W23/A188/B73"
                     /db_xref="taxon:4577"
                     /clone_lib="1006 - RescueMu Grid G"
                     /tissue_type="leaf"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="organ: leaf; Vector: RescueMu (engineered from
                     pBlueScript backbone); Site_1: BamHI; Site_2: BglII;
                     RescueMu is a 4.9 kb, modified maize Mu transposon
                     designed to allow plasmid rescue from total genomic DNA.
                     Mu elements insert preferentially into transcription
                     units. For more information on RescueMu, go to the web
                     site 'www.zmdb.iastate.edu' and follow the links for
                     'RescueMu.' Grid G was grown at Stanford in 2000. DNA was
                     extracted from leaf punches, double digested using BamHI
                     and BglII, and ligated to form circular plasmids. DH10B
                     cells were transformed and then screened on LB plates with
                     ampicillin."
BASE COUNT       63 a     64 c     81 g     91 t      1 others
ORIGIN

  Query Match           79.0%;  Score 15.8;  DB 17;  Length 300;
  Best Local Similarity   89.5%;  Pred. No. 8.1e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACC 19
           |||||||||| |||| |||
Db     73 GATCTGAATCTGACCCACC 91


RESULT 14
BM443097/c
LOCUS       BM443097                 302 bp    mRNA    linear   EST 23-JUL-2002
DEFINITION  EBro02_SQ002_I18_R root, 3 week, hydroponic grown, low nitrogen, cv
            Optic, EBro02 Hordeum vulgare cDNA clone EBro02_SQ002_I18 5', mRNA
            sequence.
ACCESSION   BM443097
VERSION     BM443097.2  GI:21947759
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 302)
  AUTHORS   Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardle,L.,
            Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
  TITLE     Development of Barley Transcriptome Resources
  JOURNAL   Unpublished (2001)
COMMENT     On Feb 1, 2002 this sequence version replaced gi:18473872.
            Contact: Waugh R, Marshall DF
            Genome Dynamics/Computational Biology
            Scottish Crop Research Institute
            Invergowrie, Dundee, DD2 5DA, Scotland, UK
            Tel: 00 44 1382 562731
            Fax: 00 44 1382 562426
            Email: est@scri.sari.ac.uk
            All sequence has a Phred quality score of 20 or over
            Seq primer: M13 reverse.
FEATURES             Location/Qualifiers
     source          1..302
                     /organism="Hordeum vulgare"
                     /cultivar="Optic"
                     /db_xref="taxon:4513"
                     /clone="EBro02_SQ002_I18"
                     /clone_lib="root, 3 week, hydroponic grown, low nitrogen,
                     cv Optic, EBro02"
                     /tissue_type="root"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
                     Non-normalised library, directionally cloned into pSPORT1.
                     Derived from roots of 3 week old Nitrogen stressed barley
                     plants. Developed as part of the barley transcriptome
                     resources of BBSRC/SEERAD funded cereal IGF (Investigating
                     Gene Function) project."
BASE COUNT       54 a     72 c     92 g     84 t
ORIGIN

  Query Match           79.0%;  Score 15.8;  DB 13;  Length 302;
  Best Local Similarity   89.5%;  Pred. No. 8.1e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACC 19
           |||| |||||||||| ||||
Db     253 GATCCGAATCCGACAAACC 235


RESULT 15
BB297513
LOCUS       BB297513                 328 bp    mRNA    linear   EST 10-JUL-2000
DEFINITION  BB297513 RIKEN full-length enriched, 9.5 days embryo parthenogenote
            Mus musculus cDNA clone B130055M05 3', mRNA sequence.
ACCESSION   BB297513
VERSION     BB297513.1  GI:8998218
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 328)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
```

```
                , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
                ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
                , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
                ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,T.
                , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
                ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE         RIKEN Mouse ESTs (Kono,H., et al.)
  JOURNAL       Unpublished (2000)
COMMENT         Contact: Yoshihide Hayashizaki
                Laboratory for Genome Exploration Research Group, RIKEN Genomic
                Sciences Center(GSC), Yokohama Institute
                The Institute of Physical and Chemical Research (RIKEN)
                1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
                Tel: 81-45-503-9222
                Fax: 81-45-503-9216
                Email: genome-res@gsc.riken.go.jp,
                URL:http://genome.gsc.riken.go.jp/
                Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
                ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
                Thermostabilization and thermoactivation of thermolabile enzymes by
                trehalose and its application for the synthesis of full length
                cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
                Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
                Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
                ,Y. and Hayashizaki,Y.
                Automated filtration-based high-throughput plasmid preparation
                system. Genome Res. 9 (5), 463-470 (1999)
                Carninci,P. and Hayashizaki,Y.
                High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
                19-44 (1999)
                Please visit our web site (http://genome.rtc.riken.go.jp) for
                further details.
FEATURES                Location/Qualifiers
     source             1..328
                        /organism="Mus musculus"
                        /db_xref="taxon:10090"
                        /clone="B130055M05"
                        /clone_lib="RIKEN full-length enriched, 9.5 days embryo
                        parthenogenote"
                        /tissue_type="parthenogenote"
                        /dev_stage="9.5 days embryo"
                        /lab_host="DH10B"
                        /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                        prepared and sequenced in Mouse Genome Encyclopedia
                        Project of Genome Exploration Research Group in Riken
                        Genomic Sciences Center and Genome Science Laboratory in
                        RIKEN. Division of Experimental Animal Research in Riken
                        contributed to prepare mouse tissues. 1st strand cDNA was
                        primed with a primer [5'
                        GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTVN 3']. cDNA was
                        prepared by using trehalose thermo-activated reverse
                        transcriptase and subsequently enriched for full-length by
                        cap-trapper. cDNA went through one round of subtraction to
                        Rot = 229.0 Second strand cDNA was prepared with the
                        primer adapter of sequence [5'
                        GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                        was cleaved with XhoI and BamHI. Vector: a modified
                        pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT        96 a     87 c     57 g     88 t
ORIGIN

  Query Match           79.0%;  Score 15.8;  DB 10;  Length 328;
  Best Local Similarity  89.5%;  Pred. No. 8.3e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GATCTGAATCCGACCAACC 19
           ||||||||| || ||||
Db      25 GATCTGAATCTGAGCAACC 43


Search completed: May 12, 2003, 02:35:23
Job time : 249.143 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic ~nucleic search, using sw model

Run on: May 12, 2003, 01:02:58 ; Search Time 99 Seconds
(without alignments)
5879.355 Million cell updates/sec

Title: US-09-980-277-3
Perfect score: 20
Sequence: 1 atattaacgcccacgctctc 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c	20	100.0	121	3	SCMRPTAA	M61098 S.mansoni t
c 2	17.4	87.0	146970	2	AC094913	AC094913 Rattus no
c 3	16.8	84.0	1356	3	AY010508	AY010508 Scaliola
4	16.8	84.0	100685	4	AC023754	AC023754 Arabidops
5	16.8	84.0	106760	9	HUAC002990	AC002990 Human Chr
6	16.8	84.0	119535	2	AC129382	AC129382 Rattus no
7	16.8	84.0	168949	9	AC046134	AC046134 Homo sapi
8	16.8	84.0	184321	2	AC025890	AC025890 Homo sapi
c 9	16.4	82.0	3226	9	AK027375	AK027375 Homo sapi
c 10	16.4	82.0	4695	3	AF387518	AF387518 Drosophil
c 11	16.4	82.0	5878	3	AY121666	AY121666 Drosophil
c 12	16.4	82.0	14086	1	AE009916	AE009916 Pyrobacul
c 13	16.4	82.0	40123	2	AC014130	AC014130 Drosophil
14	16.4	82.0	65961	2	AC016523	AC016523 Drosophil
c 15	16.4	82.0	80866	2	AC012649	AC012649 Drosophil
16	16.4	82.0	109155	8	ATT20K12	AL137898 Arabidops
17	16.4	82.0	112314	9	AC004002	AC004002 Human BAC
c 18	16.4	82.0	154411	9	AC073909	AC073909 Homo sapi
19	16.4	82.0	163307	3	AC011703	AC011703 Drosophil
20	16.4	82.0	167688	3	AC011702	AC011702 Drosophil
21	16.4	82.0	173281	3	AC012161	AC012161 Drosophil
22	16.4	82.0	300994	3	AE003506	AE003506 Drosophil
c 23	16.4	82.0	305124	2	AC073858	AC073858 Homo sapi
24	16	80.0	591	3	DPCH3C1	X92196 D.polychroa
c 25	16	80.0	155934	2	AC116613	AC116613 Homo sapi
c 26	16	80.0	220962	2	AC126272	AC126272 Mus muscu
27	15.8	79.0	678	9	HSA340686	AJ340686 Homo sapi
28	15.8	79.0	1807	9	AF064769	AF064769 Homo sapi
29	15.8	79.0	4944	3	DMRNPOL2	X05709 Drosophila
30	15.8	79.0	5482	10	RNRJG9	X17215 Rat RJG-9 g
c 31	15.8	79.0	6075	6	AX346453	AX346453 Sequence
c 32	15.8	79.0	6207	6	AX339169	AX339169 Sequence
c 33	15.8	79.0	6922	6	AX251287	AX251287 Sequence
c 34	15.8	79.0	6922	6	AX252032	AX252032 Sequence
c 35	15.8	79.0	6922	6	AX344406	AX344406 Sequence
c 36	15.8	79.0	6922	6	AX348811	AX348811 Sequence
c 37	15.8	79.0	7631	6	AX345762	AX345762 Sequence
c 38	15.8	79.0	8076	6	AX348340	AX348340 Sequence
c 39	15.8	79.0	9543	6	AX344641	AX344641 Sequence
40	15.8	79.0	11844	9	AL391691	AL391691 Human DNA
41	15.8	79.0	11845	1	AE011176	AE011176 Methanosa
42	15.8	79.0	14344	5	AF124048	AF124048 Fugu rubr
c 43	15.8	79.0	14816	10	MMU276492	AJ276492 Mus muscu
44	15.8	79.0	16522	2	AC013170	AC013170 Drosophil
c 45	15.8	79.0	19459	6	AX344209	AX344209 Sequence

ALIGNMENTS

RESULT 1
SCMRPTAA/c
LOCUS SCMRPTAA 121 bp DNA linear INV 26-APR-1993
DEFINITION S.mansoni tandem repeat units.
ACCESSION M61098
VERSION M61098.1 GI:161064
KEYWORDS .
SOURCE S.mansoni (strain Egyptian) DNA, clone pSm1-7.
ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE 1 (bases 1 to 121)
AUTHORS Hamburger,J., Turetski,T., Kapeller,I. and Deresiewicz,R.
TITLE Highly repeated short DNA sequences in the genome of Schistosoma mansoni recognized by a species-specific probe

```
  JOURNAL   Mol. Biochem. Parasitol. 44 (1), 73-80 (1991)
  MEDLINE   91187067
    PUBMED  2011155
FEATURES             Location/Qualifiers
     source          1. .121
                     /organism="Schistosoma mansoni"
                     /db_xref="taxon:6183"
BASE COUNT         36 a    20 c    26 g    39 t
ORIGIN

  Query Match             100.0%;  Score 20;  DB 3;  Length 121;
  Best Local Similarity   100.0%;  Pred. No. 9.1;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCTC 20
           ||||||||||||||||||||
Db     110 ATATTAACGCCCACGCTCTC 91


RESULT 2
AC094913/c
LOCUS       AC094913               146970 bp    DNA     linear   HTG 10-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-6C20, *** SEQUENCING IN PROGRESS ***,
            59 unordered pieces.
ACCESSION   AC094913
VERSION     AC094913.3  GI:21716171
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 146970)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Liew,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Rea,Y.,
            Rives,M., Rojas,A., Robubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Soderqren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wieczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.

  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 146970)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 146970)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 9, 2002 this sequence version replaced gi:17941710.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: CBRX
                Center clone name: CH230-6C20
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 92271 bases at least Q40
                Consensus quality: 97706 bases at least Q30
                Consensus quality: 102314 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/Goss/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 59 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    1179: contig of 1179 bp in length
            *    1180    1279: gap of unknown length
            *    1280    2496: contig of 1217 bp in length
            *    2497    2596: gap of unknown length
            *    2597    4170: contig of 1574 bp in length
            *    4171    4270: gap of unknown length
            *    4271    5575: contig of 1305 bp in length
            *    5576    5675: gap of unknown length
            *    5676    7019: contig of 1344 bp in length
            *    7020    7119: gap of unknown length
            *    7120    8691: contig of 1572 bp in length
            *    8692    8791: gap of unknown length
            *    8792   10183: contig of 1392 bp in length
            *   10184   10283: gap of unknown length
            *   10284   11471: contig of 1188 bp in length
            *   11472   11571: gap of unknown length
            *   11572   12723: contig of 1152 bp in length
            *   12724   12823: gap of unknown length
            *   12824   14403: contig of 1580 bp in length
            *   14404   14503: gap of unknown length
            *   14504   15527: contig of 1024 bp in length
            *   15528   15627: gap of unknown length
            *   15628   17611: contig of 1984 bp in length
            *   17612   17711: gap of unknown length
            *   17712   19429: contig of 1718 bp in length
            *   19430   19529: gap of unknown length
            *   19530   21227: contig of 1698 bp in length
            *   21228   21327: gap of unknown length
            *   21328   22376: contig of 1049 bp in length
            *   22377   22476: gap of unknown length
            *   22477   24009: contig of 1533 bp in length
            *   24010   24109: gap of unknown length
```

```
*    24110    26423: contig of 2314 bp in length
*    26424    26523: gap of unknown length
*    26524    27957: contig of 1434 bp in length
*    27958    28057: gap of unknown length
*    28058    29638: contig of 1581 bp in length
*    29639    29738: gap of unknown length
*    29739    31846: contig of 2108 bp in length
*    31847    31946: gap of unknown length
*    31947    34142: contig of 2196 bp in length
*    34143    34242: gap of unknown length
*    34243    36376: contig of 2134 bp in length
*    36377    36476: gap of unknown length
*    36477    38561: contig of 2085 bp in length
*    38562    38661: gap of unknown length
*    38662    39903: contig of 1242 bp in length
*    39904    40003: gap of unknown length
*    40004    41958: contig of 1955 bp in length
*    41959    42058: gap of unknown length
*    42059    43625: contig of 1567 bp in length
*    43626    43725: gap of unknown length
*    43726    45011: contig of 1286 bp in length
*    45012    45111: gap of unknown length
*    45112    47277: contig of 2166 bp in length
*    47278    47377: gap of unknown length
*    47378    49421: contig of 2044 bp in length
*    49422    49521: gap of unknown length
*    49522    51042: contig of 1521 bp in length
*    51043    51142: gap of unknown length
*    51143    53482: contig of 2340 bp in length
*    53483    53582: gap of unknown length
*    53583    55075: contig of 1493 bp in length
*    55076    55175: gap of unknown length
*    55176    57330: contig of 2155 bp in length
*    57331    57430: gap of unknown length
*    57431    60307: contig of 2877 bp in length
*    60308    60407: gap of unknown length
*    60408    62705: contig of 2298 bp in length
*    62706    62805: gap of unknown length
*    62806    64827: contig of 2022 bp in length
*    64828    64927: gap of unknown length
*    64928    67634: contig of 2707 bp in length
*    67635    67734: gap of unknown length
*    67735    69903: contig of 2169 bp in length
*    69904    70003: gap of unknown length
*    70004    72964: contig of 2961 bp in length
*    72965    73064: gap of unknown length
*    73065    75860: contig of 2796 bp in length
*    75861    75960: gap of unknown length
*    75961    77903: contig of 1943 bp in length
*    77904    78003: gap of unknown length
*    78004    80743: contig of 2740 bp in length
*    80744    80843: gap of unknown length
*    80844    83862: contig of 3019 bp in length
*    83863    83962: gap of unknown length
*    83963    86802: contig of 2840 bp in length
*    86803    86902: gap of unknown length
*    86903    90056: contig of 3154 bp in length
*    90057    90156: gap of unknown length
*    90157    92841: contig of 2685 bp in length
*    92842    92941: gap of unknown length
*    92942    94466: contig of 1525 bp in length
*    94467    94566: gap of unknown length
*    94567    97310: contig of 2744 bp in length
*    97311    97410: gap of unknown length
*    97411    99793: contig of 2383 bp in length
*    99794    99893: gap of unknown length
*    99894   103484: contig of 3591 bp in length
*   103485   103584: gap of unknown length
*   103585   106611: contig of 3027 bp in length
*   106612   106711: gap of unknown length
*   106712   110006: contig of 3295 bp in length
*   110007   110106: gap of unknown length
*   110107   113951: contig of 3845 bp in length
*   113952   114051: gap of unknown length
*   114052   118104: contig of 4053 bp in length

  Query Match              87.0%;  Score 17.4;  DB 2;  Length 146970;
  Best Local Similarity    94.7%;  Pred. No. 2e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 TATTAACGCCCACGCTCTC 20
           ||||||||||||||||| ||
Db  139691 TATTAACGCCCACGCTTTC 139673


RESULT 3
AY010508/c
LOCUS       AY010508                1356 bp    DNA     linear   INV 15-MAR-2002
DEFINITION  Scaliola sp. tRNA-Thr, tRNA-Gly, and tRNA-Val genes, complete
            sequence; and 16S ribosomal RNA gene, partial sequence;
            mitochondrial genes for mitochondrial products.
ACCESSION   AY010508
VERSION     AY010508.1  GI:16554857
KEYWORDS
SOURCE      Scaliola sp.
  ORGANISM  Mitochondrion Scaliola sp.
            Eukaryota; Metazoa; Mollusca; Gastropoda; Caenogastropoda;
            Neotaenioglossa; Cerithioidea; Scaliolidae; Scaliola.
REFERENCE   1  (bases 1 to 1356)
  AUTHORS   Lydeard,C., Holznagel,W.E., Glaubrecht,M. and Ponder,W.F.
  TITLE     Molecular phylogeny of a circum-global, diverse gastropod
            superfamily (Cerithioidea: Mollusca: Caenogastropoda): pushing the
            deepest phylogenetic limits of mitochondrial LSU rDNA sequences
  JOURNAL   Mol. Phylogenet. Evol. 22 (3), 399-406 (2002)
  MEDLINE   21881977
   PUBMED   11884164
REFERENCE   2  (bases 1 to 1356)
  AUTHORS   Lydeard,C., Holznagel,W.E., Glaubrecht,M. and Ponder,W.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-OCT-2000) Biodiversity and Systematics, Department of
            Biological Sciences, The University of Alabama, Box 870345,
            Tuscaloosa, AL 35487-0345, USA
FEATURES             Location/Qualifiers
     source          1..1356
                     /organism="Scaliola sp."
                     /organelle="mitochondrion"
                     /db_xref="taxon:145863"
     tRNA            1..64
                     /product="tRNA-Thr"
     tRNA            65..128
                     /product="tRNA-Gly"
     tRNA            129..195
                     /product="tRNA-Val"
     rRNA            196..>1356
                     /product="16S ribosomal RNA"
BASE COUNT      517 a    188 c    218 g    433 t
ORIGIN

  Query Match              84.0%;  Score 16.8;  DB 3;  Length 1356;
  Best Local Similarity    90.0%;  Pred. No. 4.5e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCTC 20
           |||||||||||||||| | ||
Db    1125 ATATTAACGCCCACGGTTTC 1106


RESULT 4
AC023754
LOCUS       AC023754               100685 bp    DNA     linear   PLN 13-SEP-2000
DEFINITION  Arabidopsis thaliana chromosome I BAC F18I16 genomic sequence,
            complete sequence.
ACCESSION   AC023754
VERSION     AC023754.3  GI:7212003
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana.
```

```
ORGANISM  Arabidopsis thaliana
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1  (bases 1 to 100685)
  AUTHORS Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
          Altafi,H., Nguyen,M., Lam,B., Southwick,A., Miranda,M., Brooks,S.,
          Buehler,E., Chao,Q., Chin,C., Chiou,J., Choi,E., Gonzalez,A.,
          Howng,B., Johnson-Hopson,C., Khan,S., Kim,C., Koo,T., Lee,J.M.,
          Lenz,C., Liu,A., Liu,S., Mukharsky,N., Pham,P., Sakano,H.,
          Shinn,P., Toriumi,M., Vaysberg,M., Yu,G., Ecker,J., Theologis,A.
          and Davis,R.W.
  JOURNAL Unpublished
REFERENCE 2  (bases 1 to 100685)
  AUTHORS Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
          Altafi,H., Nguyen,M., Lam,B., Southwick,A., Bei,Q., Buehler,E.,
          Chin,C., Chiou,J., Choi,E., Dunn,P., Gonzalez,A., Howng,B., Kim,C.,
          Koo,T., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,
          Mukharsky,N., Pham,P., Sakano,H., Schwartz,J., Shinn,P.,
          Thaveri,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J.,
          Theologis,A. and Davis,R.W.
  TITLE   Direct Submission
  JOURNAL Submitted (17-FEB-2000) DNA Sequencing and Technology Center,
          Stanford University, 855 California Avenue, Palo Alto, CA 94304,
          USA
REFERENCE 3  (bases 1 to 100685)
  AUTHORS Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
          Altafi,H., Nguyen,M., Lam,B., Southwick,A., Bei,Q., Buehler,E.,
          Chin,C., Chiou,J., Choi,E., Dunn,P., Gonzalez,A., Howng,B., Kim,C.,
          Koo,T., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,
          Mukharsky,N., Pham,P., Sakano,H., Schwartz,J., Shinn,P.,
          Thaveri,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J.,
          Theologis,A. and Davis,R.W.
  TITLE   Direct Submission
  JOURNAL Submitted (09-MAR-2000) DNA Sequencing and Technology Center,
          Stanford University, 855 California Avenue, Palo Alto, CA 94304,
          USA
REFERENCE 4  (bases 1 to 100685)
  AUTHORS Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
          Altafi,H., Nguyen,M., Lam,B., Southwick,A., Ecker,J., Theologis,A.
          and Davis,R.W.
  TITLE   Direct Submission
  JOURNAL Submitted (13-SEP-2000) DNA Sequencing and Technology Center,
          Stanford University, 855 California Avenue, Palo Alto, CA 94304,
          USA
COMMENT   On Mar 9, 2000 this sequence version replaced gi:7143418.
          Bases 1-32,040 of IGF clone F1B16 overlap with bases 97,508-129,547
          of IGF clone F10A5, gb|AC006434.
           e-mail for  correspondence: arab@sequence.stanford.edu
          Genes with similarity to  proteins in the databases are named
          'putative', '-like' or 'similar to'.  Genes that have EST
          similarity but no significant protein similarity are  described as
          'unknown proteins'.  Genes that are annotated based only on  gene
          prediction software are described as 'hypothetical proteins'.  The
           gene  prediction programs used to predict genes include:  Grail
          (Informatics  Group,  Oak Ridge National Laboratory,
          http://compbio.ornl.gov/section/index.html),  GENSCAN (Chris Burge,
           http://genes.mit.edu/GENSCAN.), FEXA (Victor Solovyev,
          http://genomic.sanger.ac.uk/gf/gf.shtml),  and NetPlantGene (S.M.
          Hebsgaard,   et al., CBS, Technical University of
          Denmark,http://www.cbs.dtu.dk/NetPlantGene.html).
FEATURES             Location/Qualifiers
     source          1. .100685
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="I"
                     /clone="F1B16"
     misc_feature    1. .32040
                     /note="overlap with bases 97,508-129,547 of IGF clone
                     F10A5, gb|AC006434.  See GenBank record for BAC clone
                     F10A5 for annotation in this region."
     gene            complement(32986. .33822)
                     /gene="F1B16.1"
     CDS             complement(32986. .33822)
                     /gene="F1B16.1"
                     /note="Unknown protein; unknown protein"
                     /codon_start=1
                     /protein_id="AAG13069.1"
                     /db_xref="GI:10120444"
                     /translation="MALPHVVSSPSSTSLSPSFKSKPPFHSLRSLSPTVDNRRRCKLN
                     SSFRCSSSSFSEKHHNAKSPKSDDIVELPLFPLPLVLFPGAILPLQIFEFRYRIMMHT
                     LLQSDLRFGVVYSDSVSGSAAEVGCVGEVVKHERLVDDRFFLVCKGQERFRVTNVVRT
                     KPYLVGEVTWLEDRPSGEENLDSLANEVEVLMKEVIRLSNRLNGKAEKEVQDLRRNQF
                     PTPFSFFVGSTFEGAPREQQALLELEDTAARLKRERETLRNTLNYLTAASAVKDVFPS
                     SS"
     gene            43756. .46560
                     /gene="F1B16.2"
     CDS             join(43756. .44347,44507. .44562,44700. .44966,45493. .45755,
                     46200. .46560)
                     /gene="F1B16.2"
                     /note="Similar to cytokinin oxidase"
                     /codon_start=1
                     /protein_id="AAG13068.1"
                     /db_xref="GI:10120443"
                     /translation="MTSSFLLLTFAICKLIIAVGLNVGPSELLRIGAIDVDGHFTVHP
                     SDLASVSSDFGMLKSPEEPLAVLHPSSAEDVARLVRTAYGSATAFPVSARGHGHSING
                     QAAAGRNGVVVEMNHGVTGTPKPLVRPDEMYVDVWGGELWVDVLKKTLEHGLAPKSWT
                     DYLYLTVGGTLSNAGISGQAFHHGPQISNVLELDVVTGQFGIITRARISLEPAPQRVR
                     WIRVLYSSFKVFTEDQEYLISMHGQLKFDYVEGFVIVDEGLVNNWRSSFFSPRNPVKI
                     SSVSSNGSVLYCLEITKNYHDSDSEIVDQEVEILMKKLNFIPTSVFTTDLQYVDFLDR
                     VHKAELKLRSKNLWEVPHPWLNLFVPKSRISDFDKGVFKGILGNKTSGPILIYPMNKD
                     KWDERSSAVTPDEEVFYLVALLRSALTDGEETQKLEYLKDQNRRILEFCEQAKINVKQ
                     YLPHHATQEEWVAHFGDKWDRFRSLKAEFDPRHILATGQRIFQNPSLSLFPPSSSSSS
                     AASW"
     gene            complement(46848. .48185)
                     /gene="F1B16.3"
     CDS             complement(join(46848. .46943,47040. .47128,47474. .47537,
                     47773. .47875,47963. .48066,48156. .48185))
                     /gene="F1B16.3"
                     /codon_start=1
                     /product="Putative ubiquitin-conjugating enzyme"
                     /protein_id="AAG13066.1"
                     /db_xref="GI:10120441"
                     /translation="MSSSGAPSRKTLSKIATNRLQKELVEWQMNPPTGFKHKVTDNLQ
                     RWIIEVIGAPGTLYANDTYQLQVDFPEHYPMESPQVIFLHPAPLHPHIYSNGHICLDI
                     LYDSWSPAMTVSSICISILSMLSSSTEKQRPTDNDRYVKNCKNGRSPKETRWWFHDDK
                     V"
     gene            52291. .53501
                     /gene="F1B16.4"
     CDS             join(52291. .52584,52917. .53501)
                     /gene="F1B16.4"
                     /note="Hypothetical protein"
                     /codon_start=1
                     /protein_id="AAG13063.1"
                     /db_xref="GI:10120438"
                     /translation="MEDFRVRHECSSLRGTLLDSRYAKAVQCLVEEVIDIGGREVELC
                     NNILINQLFPGRRRPGFALSSEIKSELCSSGFMSLPENHEIHIKITKLLSLLQQVEER
                     FEQYCNQLEQVISSFEEIAGEGSSKVYTGLALQAMTRHFGSLEEAIISQLNSVRRRFI
                     ISHQDVPKIISSGLSQLSLFDGNTTSSSLQRLGLVQGPQRHAWKPIRGLPETSVAILR
                     AWLFQHFLHPSVFLLLKCNNTSDESKLKCWRNNISLIAVMKTVIRMKQRSWCWRLKQD
                     FRRTKYTPLNLILLFF"
     gene            complement(54491. .56307)
                     /gene="F1B16.5"
     CDS             complement(join(54491. .55074,55197. .55351,55498. .55769,
                     55872. .56014,56253. .56307))
                     /gene="F1B16.5"
                     /note="Hypothetical protein"
                     /codon_start=1
                     /protein_id="AAG13070.1"
                     /db_xref="GI:10120445"
                     /translation="MGTDGFVVSICINSVYDSRGPLLLMELAFLLRGVGADVVWITNQ
                     KPLEDDEVVYSLEHKMLDRGVQVISAKGQKAVDTSLKADLIVLNTAVAGKWLDAVLKE
                     NVVKVLPKILWWIHEMRGHYFNADLVKHLPFVAGAMIDSHATAGYWKNRTQARLGIKM
                     PKTYVVHLGNSKELMEVAEDSVAKRVLREHVRESLGVRNEDLLFGIINSVSRGKGQDL
                     FLRAFHESLERIKEKKLQVPTMHAVVVGSDMSKQTKFETELRNFVREKKLENFVHFVN
                     KTLTVAPYIAAIDVLVQNSQARGECFGRITIEAMAFKLPVLGTAAGGTMEIVVNGTTG
```

```
                     LLHSAGKEGVIPLAKNIVKLAFQVELRLRMGKNGYERVKEMFLEHHMSHRIASVLKEV
                     LQHAKAARLR"
     gene            59918. .61713
                     /gene="FlB16.6"
     CDS             join(59918. .60655,60734. .61110,61188. .61248,61315. .61713)
                     /gene="F1B16.6"
                     /note="Similar to  homeodomain proteins"
                     /codon_start=1
                     /protein_id="AAG13065.1"
                     /db_xref="GI:10120440"
                     /translation="MAVTYPNSVGMQSLIQESITLNRQQQQQQASSSAASFSEIVS
                     GDVENNEMVFIPPSDVAVNGNVTVSSNDLSFHGGDLSLSLGNQIQSAVSVSPFQYHY
                     QNLSNQLSYNNLNPSTMSDENGKSLSVHQHHSDQILPSSVYNNNGNNGVGFYNNYRYE
                     TSGFVSSVLRSKYLKPTQQLLDEVVSVREDLKLGNKKMENDKGQDFHNGSSDNITEDD
                     KSQSQELSPSERQELQSKKSKLLDNVDEVDKRYMCYHQMEALASSFDNVTQGLGAAKP
                     YTSVALNRISRHFRCLRDATKEQIQVIRGKLGEEETSDEQGERIPRLRYLDQRLRQQR
                     ALHQQLGMVRPAWRPQRGLPENSVSILRAWLFSNPLHPYPKESKIMLSKQTGLSKNQ
                     VANWFINARVRLWKPMIEHMYKEEPGESAKLLSNSNQDTKKMQETSQLKEHDSSSSQQ
                     QNQGNNNNNIPTTSDASQNLVYPADPKPDRATTGDYDSLMNYHGFGIDDYNRIVYGLGNQ
                     QDGRYSNPHQLHDFVV"
     gene            complement(62425. .64698)
                     /gene="F1B16.7"
     CDS             complement(join(62425. .63429,63528. .63559,63863. .64074,
                     64640. .64698))
                     /gene="F1B16.7"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="AAG13067.1"
                     /db_xref="GI:10120442"
                     /translation="MPFRIRVSCFDFPRNLSLFWRRVADEIEGSPYYSPYVASRGISM
                     DKMSLSGERQTLSSGGPPPDGHLGTPASQKSATPEMSTNSKVPPSGSGSLASSDLIEV
                     KNLVDSRGIVPSAQSKPLPSTPSLPNPVSDLSSAHTRLLSPKSTPCRRARRSPGHQLF
                     RQVSDSQLLGLKSPINNYSISRGHSSPVLGTCSNDFATQSQTASSSGGWSMNTPSELV
                     AYSQRDRWSFDSEHLGSGRKLSGGSSRKFSPSPSVVDQQVCGACSKLLTERSSIATPE
                     LPIAAVLACGHVTHAECLRTMTTDIEKYDPACPICTIGEKRVAKITRKALKAEAAAKA
                     KQYKYKNKVVSYDSYGESECDEEVFQKMGRKEGKALKLEASCSSKSSSNNKSFLKWHFAS
                     ISSKWNKPSSKDSALKKGFWSHRHRNNRSSSIEVKIHTLNIQS"
     gene            complement(69063. .69566)
                     /gene="F1B16.8"
     CDS             complement(69063. .69566)
                     /gene="F1B16.8"
                     /note="similar to DNA-binding protein"
                     /codon_start=1
                     /protein_id="AAG13064.1"
                     /db_xref="GI:10120439"
                     /translation="MGDSTSGNCSSVSTPGLANSGSESDLRQRDLIDERKKREQSNR
                     ESARSRMRKQKHLDDLTAQVTHLRKENAQIVAGIAVTTQHTVTIEAENDLRAQVLE

  Query Match          84.0%;  Score 16.8;  DB 8;  Length 100685;
  Best Local Similarity   90.0%;  Pred. No. 4.3e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCTC 20
           |||||||||| || |||||
Db   84917 ATATTAACGCCAACACTCTC 84936


RESULT 5
HUAC002990
LOCUS       HUAC002990            106760 bp    DNA     linear   PRI 23-NOV-1999
DEFINITION  Human Chromosome 16 BAC clone CIT987SK-A-1000D7, complete sequence.
ACCESSION   AC002990
VERSION     AC002990.1  GI:2815547
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 106760)
  AUTHORS   Loftus,B.J., Kim,U.J., Sneddon,V.P., Kalush,F., Brandon,R.,
            Fuhrmann,J., Mason,T., Crosby,M.L., Barnstead,M., Cronin,L.,
            Deslattes Mays,A., Cao,Y., Xu,R.X., Kang,H.L., Mitchell,S.,
            Eichler,E.E., Harris,P.C., Venter,J.C. and Adams,M.D.
  TITLE     Genome duplications and other features in 12 Mb of DNA sequence
            from human chromosome 16p and 16q
  JOURNAL   Genomics 60 (3), 295-308 (1999)
  MEDLINE   99425270
   PUBMED   10493829
REFERENCE   2  (bases 1 to 106760)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Human Chromosome 16 BAC clone CIT987SK-A-1000D7
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 106760)
  AUTHORS   Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-OCT-1997) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   4  (bases 1 to 106760)
  AUTHORS   Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JAN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   5  (bases 1 to 106760)
  AUTHORS   Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JAN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   6  (bases 1 to 106760)
  AUTHORS   Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   7  (bases 1 to 106760)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Jan 28, 1998 this sequence version replaced gi:2809277.
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
            address: mdadams@tigr.org.  The orientation of the sequence is from
            SP6 end to T7 end. Genes were identified by a combination of five
            methods including: XGRAIL (available by anonymous ftp from
            arthur.epm.ornl.gov), GeneFinder (Phil Green, University of
            Washington), Genscan (Chris Burge,
            http://genomic.stanford.EDu/~chris/GENSCANW.html)searches of the
            complete sequence against a peptide database,and the Human gene
            Index database at TIGR (http://www.tigr.org/tdb/hgi/hgi.html).
            Genes without peptide homology having spliced EST hits are termed
            'Unknown gene product'. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES                Location/Qualifiers
     source             1. .106760
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="16"
                        /map="16p12"
                        /clone="A-1000D7"
     STS                43225. .43356
                        /db_xref="dbSTS:G19948"
     STS                105684. .105815
                        /db_xref="dbSTS:G19948"
BASE COUNT    28506 a  22731 c  23448 g  32075 t
ORIGIN

  Query Match          84.0%;  Score 16.8;  DB 9;  Length 106760;
  Best Local Similarity   90.0%;  Pred. No. 4.3e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCTC 20
           ||||| |||||||  ||||||
Db   73207 ATATTTACGCCCAAGCTCTC 73226


RESULT 6
```

```
AC129382
LOCUS       AC129382               119535 bp    DNA     linear   HTG 02-AUG-2002
DEFINITION  Rattus norvegicus clone CH230-91O18, *** SEQUENCING IN PROGRESS
            ***, 56 unordered pieces.
ACCESSION   AC129382
VERSION     AC129382.2  GI:22038396
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 119535)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 119535)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 119535)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Aug 1, 2002 this sequence version replaced gi:22001463.
            ------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ------------- Project Information
                Center project name: GIXG
                Center clone name: CH230-91O18
            ------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 59977 bases at least Q40
                Consensus quality: 67676 bases at least Q30
                Consensus quality: 73666 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 56 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1002: contig of 1002 bp in length
            *     1003     1102: gap of unknown length
            *     1103     2108: contig of 1006 bp in length
            *     2109     2208: gap of unknown length
            *     2209     3843: contig of 1635 bp in length
            *     3844     3943: gap of unknown length
            *     3944     5160: contig of 1217 bp in length
            *     5161     5260: gap of unknown length
            *     5261     6590: contig of 1330 bp in length
            *     6591     6690: gap of unknown length
            *     6691     7886: contig of 1196 bp in length
            *     7887     7986: gap of unknown length
            *     7987     9069: contig of 1083 bp in length
            *     9070     9169: gap of unknown length
            *     9170    10939: contig of 1770 bp in length
            *    10940    11039: gap of unknown length
            *    11040    12437: contig of 1398 bp in length
            *    12438    12537: gap of unknown length
            *    12538    13650: contig of 1113 bp in length
            *    13651    13750: gap of unknown length
            *    13751    15173: contig of 1423 bp in length
            *    15174    15273: gap of unknown length
            *    15274    16946: contig of 1673 bp in length
            *    16947    17046: gap of unknown length
            *    17047    18643: contig of 1597 bp in length
            *    18644    18743: gap of unknown length
            *    18744    20203: contig of 1460 bp in length
            *    20204    20303: gap of unknown length
            *    20304    21681: contig of 1378 bp in length
            *    21682    21781: gap of unknown length
            *    21782    23406: contig of 1625 bp in length
            *    23407    23506: gap of unknown length
            *    23507    24901: contig of 1395 bp in length
            *    24902    25001: gap of unknown length
            *    25002    26118: contig of 1117 bp in length
            *    26119    26218: gap of unknown length
            *    26219    27516: contig of 1298 bp in length
            *    27517    27616: gap of unknown length
            *    27617    29053: contig of 1437 bp in length
            *    29054    29153: gap of unknown length
            *    29154    31107: contig of 1954 bp in length
            *    31108    31207: gap of unknown length
            *    31208    33384: contig of 2177 bp in length
            *    33385    33484: gap of unknown length
            *    33485    35707: contig of 2223 bp in length
            *    35708    35807: gap of unknown length
            *    35808    37972: contig of 2165 bp in length
            *    37973    38072: gap of unknown length
            *    38073    39104: contig of 1032 bp in length
            *    39105    39204: gap of unknown length
            *    39205    40762: contig of 1558 bp in length
            *    40763    40862: gap of unknown length
```

```
            *    40863   42690: contig of 1828 bp in length
            *    42691   42790: gap of unknown length
            *    42791   44621: contig of 1831 bp in length
            *    44622   44721: gap of unknown length
            *    44722   46492: contig of 1771 bp in length
            *    46493   46592: gap of unknown length
            *    46593   48535: contig of 1943 bp in length
            *    48536   48635: gap of unknown length
            *    48636   49726: contig of 1091 bp in length
            *    49727   49826: gap of unknown length
            *    49827   51433: contig of 1607 bp in length
            *    51434   51533: gap of unknown length
            *    51534   53994: contig of 2461 bp in length
            *    53995   54094: gap of unknown length
            *    54095   56420: contig of 2326 bp in length
            *    56421   56520: gap of unknown length
            *    56521   57610: contig of 1090 bp in length
            *    57611   57710: gap of unknown length
            *    57711   59464: contig of 1754 bp in length
            *    59465   59564: gap of unknown length
            *    59565   61644: contig of 2080 bp in length
            *    61645   61744: gap of unknown length
            *    61745   63220: contig of 1476 bp in length
            *    63221   63320: gap of unknown length
            *    63321   66614: contig of 3294 bp in length
            *    66615   66714: gap of unknown length
            *    66715   68356: contig of 1642 bp in length
            *    68357   68456: gap of unknown length
            *    68457   70817: contig of 2361 bp in length
            *    70818   70917: gap of unknown length
            *    70918   72489: contig of 1572 bp in length
            *    72490   72589: gap of unknown length
            *    72590   74365: contig of 1776 bp in length
            *    74366   74465: gap of unknown length
            *    74466   77170: contig of 2705 bp in length
            *    77171   77270: gap of unknown length
            *    77271   79757: contig of 2487 bp in length
            *    79758   79857: gap of unknown length
            *    79858   83585: contig of 3728 bp in length
            *    83586   83685: gap of unknown length
            *    83686   85419: contig of 1734 bp in length
            *    85420   85519: gap of unknown length
            *    85520   87286: contig of 1767 bp in length
            *    87287   87386: gap of unknown length
            *    87387   89241: contig of 1855 bp in length
            *    89242   89341: gap of unknown length
            *    89342   91596: contig of 2255 bp in length
            *    91597   91696: gap of unknown length
            *    91697   95810: contig of 4114 bp in length
            *    95811   95910: gap of unknown length
            *    95911   99718: contig of 3808 bp in length
            *    99719   99818: gap of unknown length
            *    99819  103856: contig of 4038 bp in length
            *   103857  103956: gap of unknown length
            *   103957  109114: contig of 5158 bp in length

  Query Match             84.0%;  Score 16.8;  DB 2;  Length 119535;
  Best Local Similarity   90.0%;  Pred. No. 4.2e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCTC 20
           ||||||| |||||||||| |
Db   91492 ATATTAAAGCCCACGCTCAC 91511


RESULT 7
AC046134
LOCUS       AC046134               168949 bp    DNA     linear   PRI 15-MAY-2002
DEFINITION  Homo sapiens 3 BAC RP11-553K23 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC046134
VERSION     AC046134.8  GI:20334471
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 168949)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaris,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carroa,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,C., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Dean,A.L., Ding,Y., Dinh,H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
            Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N.,
            Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
            Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
            Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
            Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
            Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,
            Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
            Tabor,P., Temerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
            Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
            Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
            Warren,R., Washington,C., Watlington,S., Williams,G.,
            Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
            Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 168949)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-APR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 168949)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 168949)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 168949)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Apr 27, 2002 this sequence version replaced gi:19718590.
```

INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT.

```
FEATURES             Location/Qualifiers
     source          1. .168949
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-553K23"
     misc_feature    1. .1999
                     /note="overlaps bases 65964. .67962 of clone AC110716"
                     /function="clone overlap"
     repeat_region   complement(388. .505)
                     /rpt_family="L2"
     repeat_region   1269. .1360
                     /rpt_family="MIR"
     repeat_region   1942. .1985
                     /rpt_family="(TG)n"
     repeat_region   2142. .2297
                     /rpt_family="MIR"
     repeat_region   3303. .3347
                     /rpt_family="CT-rich"
     repeat_region   complement(3360. .3455)
                     /rpt_family="L2"
     repeat_region   complement(6577. .6808)
                     /rpt_family="LTR16C"
     repeat_region   complement(7615. .7924)
                     /rpt_family="AluJb"
     repeat_region   8790. .9226
                     /rpt_family="L2"
     repeat_region   complement(9397. .9496)
                     /rpt_family="MLT1G1"
     repeat_region   9550. .10657
                     /rpt_family="L1MA4"
     repeat_region   complement(10685. .11011)
                     /rpt_family="MLT1G1"
     repeat_region   11358. .11391
                     /rpt_family="(TC)n"
     repeat_region   11542. .11710
                     /rpt_family="L2"
     repeat_region   complement(12392. .12606)
                     /rpt_family="MLT1K"
     repeat_region   complement(13485. .13607)
                     /rpt_family="MIR"
     repeat_region   13574. .13648
                     /rpt_family="MIR"
     repeat_region   13659. .14141
                     /rpt_family="MLT1D"
     repeat_region   14143. .14254
                     /rpt_family="MIR"
     repeat_region   complement(14657. .14871)
                     /rpt_family="MIR"
     repeat_region   15168. .15431
                     /rpt_family="L1MC5"
     repeat_region   complement(15432. .15730)
                     /rpt_family="AluJo"
     repeat_region   15731. .15760
                     /rpt_family="L1MC5"
     repeat_region   complement(17092. .17266)
                     /rpt_family="MER5A"
     repeat_region   17276. .17445
                     /rpt_family="MER5B"
     repeat_region   17452. .17662
                     /rpt_family="MER113"
     repeat_region   17942. .18329
                     /rpt_family="L1ME2"
     repeat_region   complement(20312. .20585)
                     /rpt_family="AluJo"
     repeat_region   20615. .20714
                     /rpt_family="L2"
     repeat_region   21914. .22316
                     /rpt_family="L2"
     repeat_region   24589. .24730
                     /rpt_family="MIR"
     repeat_region   26112. .26136
                     /rpt_family="AT_rich"
     repeat_region   26537. .26646
                     /rpt_family="MIR"
     repeat_region   27763. .27794
                     /rpt_family="(TTTG)n"
     repeat_region   complement(28315. .28354)
                     /rpt_family="L1PA10"
     repeat_region   complement(28725. .29083)
                     /rpt_family="THE1C"
     repeat_region   complement(29224. .29691)
                     /rpt_family="MLT1D"

  Query Match             84.0%;  Score 16.8;  DB 9;  Length 168949;
  Best Local Similarity   90.0%;  Pred. No. 4.2e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCTC 20
           ||||||| ||||| |||||
Db    1233 ATATTAACTCCCACCCTCTC 1252


RESULT 8
AC025890
LOCUS       AC025890               184321 bp    DNA     linear   HTG 24-AUG-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-553K23 map 3, WORKING DRAFT
            SEQUENCE, 4 unordered pieces.
ACCESSION   AC025890
VERSION     AC025890.3  GI:8077134
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 184321)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 3, clone RP11-553K23
```

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 184321)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRoccue,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mizda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triglio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (16-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 184321)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRoccue,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mizda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triglio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (24-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On May 25, 2000 this sequence version replaced gi:7387411.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
Center project name: L8244
Center clone name: 553_K_23
-------------- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap version 0.960731
Consensus quality: 181869 bases at least Q40
Consensus quality: 183483 bases at least Q30
Consensus quality: 183879 bases at least Q20
Insert size: 215000; agarose-fp
Insert size: 184021; sum-of-contigs
Quality coverage: 4.8 in Q20 bases; agarose-fp
Quality coverage: 5.6 in Q20 bases; sum-of-contigs
--------------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 18607: contig of 18607 bp in length
* 18608 18707: gap of 100 bp
* 18708 53507: contig of 34800 bp in length
* 53508 53607: gap of 100 bp
* 53608 97651: contig of 44044 bp in length
* 97652 97751: gap of 100 bp
* 97752 184321: contig of 86570 bp in length.

FEATURES Location/Qualifiers
source 1..184321
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="RP11-553K23"
/clone_lib="RPCI-11 Human Male BAC"
misc_feature 1..18607
/note="assembly_fragment
clone_end:SP6
vector_side:left"
misc_feature 18708..53507
/note="assembly_fragment"
misc_feature 53608..97651
/note="assembly_fragment"
misc_feature 97752..184321
/note="assembly_fragment
clone_end:T7
vector_side:left"

BASE COUNT 51979 a 38476 c 39574 g 53992 t 300 others
ORIGIN

Query Match 84.0%; Score 16.8; DB 2; Length 184321;
Best Local Similarity 90.0%; Pred. No. 4.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATATTAACGCCACGCTCTC 20
|||||||| |||||||||||
Db 1239 ATATTAACTCCCACCCTCTC 1258

RESULT 9
AK027375/c
LOCUS AK027375 3226 bp mRNA linear PRI 01-AUG-2002
DEFINITION Homo sapiens cDNA FLJ14469 fis, clone MAMMA1000897, weakly similar to INTER-ALPHA-TRYPSIN INHIBITOR HEAVY CHAIN H3 PRECURSOR.
ACCESSION AK027375
VERSION AK027375.1 GI:14042008
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens mammary gland cDNA to mRNA, clone_lib:MAMMA1 clone:MAMMA1000897.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H., Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,K., Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Nasuho,Y., Hinomiya,K. and Tsuyanagi,T.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3226)

```
 AUTHORS   Isogai,T. and Otsuki,T.
 TITLE     Direct Submission
 JOURNAL   Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,
           Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
           (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT    NEDO human cDNA sequencing project supported by Ministry of
           Economy, Trade and Industry of Japan; cDNA full insert sequencing:
           Research Association for Biotechnology; cDNA library construction,
           5'- & 3'-end one pass sequencing and clone selection: Helix
           Research Institute (supported by Japan Key Technology Center etc.)
           and Department of Virology, Institute of Medical Science,
           University of Tokyo.
FEATURES             Location/Qualifiers
     source          1. .3226
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MAMMA1000897"
                     /tissue_type="mammary gland"
                     /clone_lib="MAMMA1"
                     /note="cloning vector: pME18SFL3"
     CDS             44. .2872
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAB55070.1"
                     /db_xref="GI:14042009"
                     /translation="MLLLLGLCLGLSLCVGSQEEAQSWGHSSEQDGLRVPRQVRLLQR
                     LKTKPLMTEFSVKSTIISRYAFTTVSCRMLNRASEDQDIEFQMQIPAAAFITNFTMLI
                     GDKVYQGEITEREKKSGDRVKEKRNKTTEENGEKGTEIFRASAVIPSKDKAAFFLSYE
                     ELLQRRLGKYEHSISVRPQQLSGRLSVGVNILESAGIASLEVLPLHNSRQRGSGRGED
                     DSGPPPSTVINQNETFANIIFKPTVVQQARIAQNGILGDFIIRYDVNREQSIGDIQVL
                     NGYFVHYFAPKDLPPLPKNVVFVLDSSASMVGTKLRQTKDALFTILHDLRPQDRFSII
                     GFPNRIKVWKDHLISVTPDSIRDGKVYIHHMSPTGGTDINGALQRAIRLLNKYVAHSG
                     IGDRSVSLIVFLTDGKPTVGETHTLKILNNTREAARGQVCIFTIGIGNDVDFRLLEKL
                     SLENCGLTRRVHEEEDAGSQLIGFYDEIRTPLLSDIRIDYPPSSVVQATKTLFPNYFN
                     GSEIIIAGKLVDRKLDHLHVEVTASNSKKFIILKTDVPVRPQKAGKDVTGSPRPGGDG
                     EGDPNHIERLWSYLTTKELLSSWLQSDDEPEKERLRQRAQALAVSYRFLTPFTSMKLR
                     GPVPRMDGLEEAHGMSAAMGPEPVVQSVRGAGTQPGPLLKKPYQPRIKISKTSVDGDP
                     HFVVDFPLSRLTVCFNIDGQPGDILRLVSDHRDSGVTVNGELIGAPAPPNGHKKQRTY
                     LRTITILINKPERSYLEITPSRVILDGGDRLVLPCNQSVVVGSWGLEVSVSANANVTV
                     TIQGSIAFVILIHLYKKPAPFQRHHLGFYIANSEGLSSNCHGLLGQFLNQDARLTEDP
                     AGPSQNLTHPLLLQVGEGPEAVLTVKGHQVPVVWKQRKIYNGEEQIDCWFARNNAAKL
                     IDGEYKDYLASHPFDTGMTLGQGMSREL"
BASE COUNT      785 a    894 c    902 g    645 t
ORIGIN

  Query Match             82.0%;  Score 16.4;  DB 9;  Length 3226;
  Best Local Similarity   94.4%;  Pred. No. 7.3e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 ATATTAACGCCCACGCTC 18
            ||||| ||||||||||||
Db      615 ATATTCACGCCCACGCTC 598


RESULT 10
AF387518/c
LOCUS       AF387518                4695 bp    mRNA    linear   INV 29-OCT-2001
DEFINITION  Drosophila melanogaster Rho-GTPase activating protein p190 mRNA,
            complete cds.
ACCESSION   AF387518
VERSION     AF387518.1  GI:15553436
KEYWORDS    .
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 4695)
  AUTHORS   Billuart,P., Winter,C.G., Maresh,A., Zhao,X. and Luo,L.
  TITLE     Regulating axon branch stability: the role of p190 RhoGAP in
            repressing a retraction signaling pathway
  JOURNAL   Cell 107 (2), 195-207 (2001)
  MEDLINE   21526631
   PUBMED   11672527
REFERENCE   2  (bases 1 to 4695)
  AUTHORS   Billuart,P., Winter,C.G., Maresh,A., Zhao,X. and Luo,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAY-2001) Biological Sciences, Stanford University,
            385 Serra Mall, Stanford, CA 94305-5020, USA
FEATURES             Location/Qualifiers
     source          1. .4695
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
     CDS             1. .4695
                     /note="RhoGAP; includes adjacent predicted proteins CG8240
                     and CG8267: GenBank Accession Numbers AAF48749 and
                     AAF48748"
                     /codon_start=1
                     /product="Rho-GTPase activating protein p190"
                     /protein_id="AAL01872.1"
                     /db_xref="GI:15553437"
                     /translation="MRQFNISVIGLSGTEKDRGQVGVGKSCLCNRFMRPMADDYFIDH
                     ISVLSQSDFSGRIVNNDHFLYWGDVRKTTEEGVEYQFNIIEQTEFMDDSTFQAFKVGK
                     MDPYSKRCTATKVFSAEKLMYICKNQLGIEKEYEQKVMPDGRLSIDGFVVVFDVSPVP
                     NRSVEKQVEFVQNVIATILKNKKPLVLVTTKNDDAYELYVREAEKISQRKDYKSTVQL
                     IETSAHESINIDLAFLLLAQMIDKVKNRVKIISYQESAKSRKELLDTRSEAVTRLIRN
                     QITDYHVLWSQGSKMLSQYREWNEFLNIFGHEAGQKLFRRHMKKLRDDHLNKKLHQYL
                     DKFALALEYLLPDIGALNISDDDAWECARNYLQNHIEFEQYFFECPQASWTELVDMDE
                     AEDEARIPFDVLETSEAETVFRNYLNSVQQDKKKIGWKQQFKMLLEESGFVTPGKQLS
                     EVRVLFMGRECFEALSEHDCQQIYDIHQDDIIEKSKQNFVELLLEHAQYFLQFKNVDN
                     ITQEDVRQITDVIQEDSRYKMLDRLDQERRLMLVQHLRFIHCPIRDHCPFFYNCVDSL
                     IEEVLSDKSASNHKTPSGGGWKSSGSGSDRTLNLLIVGSEHLASDLLNDIRICTGSKG
                     EYIYENQTYYLNYRIANGDMEAFKAIDVYSSGLICVYSNQQSFETLKDNLERTLLCNL
                     ELEDKFENLPIVLVYQPQDLKENEVEYLRNEGMRLSEMLHCDFIDHTQNHQKYVYDIL
                     NIVILSLKLTEMKSYEPYPSNHTDLRILCCIFCGDQYDIENIVQPLVEESTLVKANEH
                     SIIVDVFIGDAKRRVEFILSSYHGTSQYRDELIHGYIYFYSTKRRSSLANLSILAAQN
                     ANIPLQIIAVTESGGVNAFFNSDICQFLITEGNAVADRFKGSFMTFSADQYVKFAFYN
                     PFLKTAWDNKYEVENLHVEESITLDSGEGTLENSVNQMPRPPPRHESYMLSNTLGTDG
                     SGSENYEMAPTRSLNSLNEERDISLDEIYDDNEKPKHLHQKWLEDKSDGRRNMNKNLI
                     WNNFSGSTHAYTTGRRHIDSNLNKIRPKGPSQTLKVGEAPSRNCPAMSSSTFTLPTQQ
                     PGKLNMKNFQLVSDAVAKMNFTGSGSGSGSGSGSGSTGLGLGLGSGSGCMGDSFLEPV
                     DKDGKRYDHAQLDGEDEDSEELAEYEQIYENEDCTESDSCASSTERRVRQQNAYYKAS
                     KKPVAAKKQKKKKVAIPVQTPRVPPFGSYVSPPEIPLHYQRMAVGGSGPGEKKKPEPC
                     VPEFMKSDKSPEYSMVPELAGAGIFGAENLPEYNMNQAKCLKDFEKLEKRRIKEETAR
                     QRKLQEKEKEQEKKLKRKLKQNAKGLVESAEAQFGKLMITSEQGEIPIFLNKCVEFIE
                     KEGLDSEGIYRVPGSRAHVDMLFQRFEEDTNTEIDALDIPVNAVATALKDFFSKRLPP
                     LFSKDIIKELEEIAGSRGVGNSKLNVEVKTDRSCRLIALKSLLQKLPPINFAILKYIF
                     QHFVHVSDNSKLNSMDSKNLAICWWPTLIPIDFTDMGHFEQLRPYLEDIVQTMIDQFP
                     YLFCGKDAFVMV"
BASE COUNT     1215 a   1212 c   1301 g    967 t
ORIGIN

  Query Match             82.0%;  Score 16.4;  DB 3;  Length 4695;
  Best Local Similarity   94.4%;  Pred. No. 7.3e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        3 ATTAACGCCCACGCTCTC 20
            |||||||||| |||||||
Db     2619 ATTAACGCCCCCGCTCTC 2602


RESULT 11
AY121666/c
LOCUS       AY121666                5878 bp    mRNA    linear   INV 18-JUN-2002
DEFINITION  Drosophila melanogaster RE10888 full insert cDNA.
ACCESSION   AY121666
VERSION     AY121666.1  GI:21464379
KEYWORDS    FLI_CDNA.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 5878)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
            Champe,M., Chavez,C., Dorsett,V., Dresnek,D., Farfan,D., Frise,E.,
```

```
                George,R., Gonzalez,M., Guarin,H., Kronmiller,B., Li,P., Liao,G.,
                Miranda,A., Mungall,C.J., Nuncio,J., Pacleb,J., Paragas,V., Park,S.,
                Patel,S., Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Rubin,G.M.
                and Celniker,S.
  TITLE         Direct Submission
  JOURNAL       Submitted (13-JUN-2002) Berkeley Drosophila Genome Project,
                Lawrence Berkeley National Laboratory, One Cyclotron Road,
                Berkeley, CA 94720, USA
COMMENT         Sequence submitted by:
                Berkeley Drosophila Genome Project
                 Lawrence Berkeley National Laboratory
                Berkeley, CA 94720
                 This clone was sequenced as part of a high-throughput process to
                sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
                Science 2000). The sequence has been subjected to integrity checks
                for sequence accuracy, presence of a polyA tail and contiguity
                within 100 kb in the genome. Thus we believe the sequence to
                reflect accurately this particular cDNA clone. However, there are
                artifacts associated with the generation of cDNA clones that may
                have not been detected in our initial analyses such as internal
                priming, priming from contaminating genomic DNA, retained introns
                due to reverse transcription of unspliced precursor RNAs, and
                reverse transcriptase errors that result in single base changes.
                For further information about this sequence, including its location
                and relationship to other sequences, please visit our Web site
                (http://fruitfly.berkeley.edu) or send email to
                cdna@fruitfly.berkeley.edu.
FEATURES                Location/Qualifiers
     source             1..5878
                        /organism="Drosophila melanogaster"
                        /strain="y; cn bw sp"
                        /db_xref="taxon:7227"
     gene               1..5878
                        /gene="RhoGAPp190"
                        /db_xref="FLYBASE:FBgn0026375"
     CDS                65..4750
                        /gene="RhoGAPp190"
                        /note="Longest ORF"
                        /codon_start=1
                        /product="RE10888p"
                        /protein_id="AAM51993.1"
                        /db_xref="GI:21464380"
                        /db_xref="FLYBASE:FBgn0026375"
                        /translation="MRQPNISVIGLSGTEKDRGQVGVGKSCLCNRFMRPMADDFYIDH
                        ISVLSQSDFSGRIVNMDHFLYWGDVRKTTREGVYEYQFNIRQTIEFMDDSTFQAFKVGK
                        NDFYSKCTKTEVFSAEKTAVICTKQGIIKYEYEQVFNPDCRLSIDGFVVFDVSNVFQ
                        NRSVEKQVFEPVQNVIATILKNKKPLVLVTTKNDDATELTVREAEKISQRKDTKSTVQL
                        IETSAHBSINIDLAFLLLAQMIDKVKNFVKIISTQESAKSRKELLDTSSEAVTRLIRN
                        QITDTHVLWSQGSKMLSQTKEWNEFLNIPGHEAGQKLFRRKMKKLRDDHLNKKLHQYL
                        DKFALALSTLLGPDIGALNISGDAWECARNTLQNHIEFEQTFFPKFPQASWTELVDMDE
                        AEDERAIPFDVLETSZAETVFRNYLNSVVQQDKKKIGWKQQFKMLLERSGFVTPGKQLS
                        EVKVLPMGRECPEALSEHDCQQIYDIHQDDIIEKSKQNFVSLLLEHAQYFLQFKYVDN
                        IPQEVFKQITNVIQNDSEYKMLDRLQRNRLMLVQNLRFIRCHIRDNCPFTFNCVDSL
                        IEHVLSDKSASNHRTPSQGGWKSSQGSDSRTLNLLIVGSEHLASDLLMDIRICTGSKG
                        EYIYRNQTTYLNYRIANGDMHAFKAIDVYSSGLICVYSNQQSFETLKDNLERTLLCNL
                        ELEEKFENLPIVLVYQPQDLKENEVEYLRNEGMKLSEMLHCDFIDHTQNHQRTVYDIL
                        NIVLSLSLNIEMKEYEPYPSNHTDLSILCCIPCGDQYDIEVTQPLVFKVKANSH
                        SIIVDVFIGDAKRRVEPTILSSYHGTSQYRDELIHGYIYPYSTKRRSSIANLSTIAAQN
                        ANIPLQIIAVPESGGVNAFFNSDICQFLITGNAVADRFKGSPMTFSADQYVKFAFYN
                        PFLATAWDNKYSVENLHVEESITYLDSGDPTLENSVNQMPRPPPRHESYMLSNTLDTDG
                        SGSENYEMATPTRSLNSLNEEDISLDEITDDNEKPKHLHQKWLEDKSDGRRNMKNNLI
                        WNNFSGSTHATTTGRRHIDSNLNKIRPKGPSQTLKVGEAPSRNCPAMSSSTPTLPTQQ
                        PGKLNWKNFQLVSDAVAKMNFTQSGSGSGSGSGSGSGSGTGLGLGSGSGCGMGDSPLEPV
                        DKDGKRTDHAQLDGEDEDSEELAETSQITPNSDCTSDSGCASSTERRVRQQNAYTKAS
                        KKPVAAKKQKKKKVAIPVQTPRVPPPGSTVSPPEIPLHTQRMAVGGSPDKPEPCVPPE
                        FMKSDKSPEYSMVPPLAGAGIPGAENLPEYNMMQAKCLKDPEKLEKRRIKEETARQRK
                        LQEEKKEQEKKLKRKLKQNAKGLVESASAQPGKLMITSSQGEIPIPLNKCVEPIEKEG
                        LDSEGIYYVPGSEAHVDMLFQRFEEDYNTEIDALDIFVNAVATALKDFFSKRLPPLFS
                        KDIIKELREIAGSRGVGNSKLNVSVKTDRSCRLIALKSLLQKLPPINFAILKYIFQHF
                        VHVSDNSKLNSMDSKNLAICWWPTLPIPIDFTDWGHFRQLRPYLEDIVQTMIDQPFYLF
                        CGSDAFVMY"
BASE COUNT      1595 a   1434 c   1581 g   1268 t
ORIGIN
```

```
  Query Match           82.0%;  Score 16.4;  DB 3;  Length 5878;
  Best Local Similarity  94.4%;  Pred. No. 7.2e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 ATTAACGCCCAGCTCTC 20
          |||||||||| |||||||
Db   2683 ATTAACGCCCCCGCTCTC 2666


RESULT 12
AE009916/c
LOCUS       AE009916              14086 bp    DNA     linear   BCT 16-JAN-2002
DEFINITION  Pyrobaculum aerophilum strain IM2 section 171 of 201 of the
            complete genome.
ACCESSION   AE009916 AE009441
VERSION     AE009916.1  GI:18161450
KEYWORDS    .
SOURCE      Pyrobaculum aerophilum.
  ORGANISM  Pyrobaculum aerophilum
            Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
            Thermoproteaceae; Pyrobaculum.
REFERENCE   1  (bases 1 to 14086)
  AUTHORS   Fitz-Gibbon,S.T., Ladner,H., Kim,U.J., Stetter,K.O., Simon,M.I. and
            Miller,J.H.
  TITLE     Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
            aerophilum
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (2), 984-989 (2002)
  PUBMED    11792869
REFERENCE   2  (bases 1 to 14086)
  AUTHORS   Fitz-Gibbon,S.T., Ladner,H., Kim,U.-J., Stetter,K.O., Simon,M.I.
            and Miller,J.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2001) Microbiology and Molecular Genetics,
            University of California, Los Angeles, 405 Hilgard Ave, Los
            Angeles, CA 90095-1489, USA
FEATURES                Location/Qualifiers
     source             1..14086
                        /organism="Pyrobaculum aerophilum"
                        /strain="IM2"
                        /db_xref="taxon:13773"
     gene               complement(118..231)
                        /gene="PAE3183"
     CDS                complement(118..231)
                        /gene="PAE3183"
                        /note="Hypothetical"
                        /codon_start=1
                        /transl_table=11
                        /product="hypothetical protein"
                        /protein_id="AAL64733.1"
                        /db_xref="GI:18161451"
                        /translation="MPLESPALLSSKKVKLGVFGTEELCGGDVSTCVGVSG"
     gene               270..322
                        /gene="PAESR34"
     misc_RNA           270..322
                        /gene="PAESR34"
                        /product="sRNA, predicted to direct ribose methylation of
                        tRNA13 (lys) at G10"
     gene               366..1457
                        /gene="PAE3184"
     CDS                366..1457
                        /gene="PAE3184"
                        /note="Protein fate; Protein modification and repair"
                        /codon_start=1
                        /transl_table=11
                        /product="pyruvate formate-lyase activating enzyme
                        homolog"
                        /protein_id="AAL64734.1"
                        /db_xref="GI:18161452"
                        /translation="MEAVLSRALPDGRVECTACARRCKLRSQQTGFCGVRTVRGGRLV
                        LDVYGLISALADPIEKKPLVHFVPGALVLSPSTFGCNWAQQTCQNFEISQRRAEGF
                        EVTPELLVKMASSYGANGITTTTTNRPTIFMEPAHDVGVLAKSKGLFNTFVTNGYMPPE
                        AVKYASEFLDAATVDFKGNADEKPMRKYIYVPDAEPIFSTLARMKRYGIWVEVTDLVV
```

```
                     PKVGDDLEKARRLVRRVIDIVGPDVPIHFLRPHPDYKMMDLPPTPVEFLEKHVEVAKK
                     EGARFAYVGNVPGHRYEHTYCPECGRVVIKRRGFDILEINLVEKGGETRCKFCGAKIP
                     IKGKIMPTWRDEFRFVYVPIQTFARWVRR"
gene                 2794. .3171
                     /gene="PAE3185"
CDS                  2794. .3171
                     /gene="PAE3185"
                     /note="Hypothetical; Conserved within genome"
                     /codon_start=1
                     /transl_table=11
                     /product="paREP9, associated with SRSRs"
                     /protein_id="AAL64735.1"
                     /db_xref="GI:18161453"
                     /translation="MYLYIGLVEIRGVRGYMWLLGRRLYLKFSWRARDTYPLGNLANP
                     LSAVSKHLFPKFVDVRAAAYAVAPALAAATVAEKCNDSPIWKVTWELKMAVEDA
                     ITYLQWIWPPTTRLPSPREPPWA"
gene                 3475. .4131
                     /gene="PAE3187"
CDS                  3475. .4131
                     /gene="PAE3187"
                     /note="Hypothetical; Conserved"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAL64736.1"
                     /db_xref="GI:18161454"
                     /translation="MQRLEKVFSSGLDELKERIDSHLRLSQDSLRILQETIASRNLDK
                     SAVDRAVRKITALERDSDEIVRDVVDRVAQGAVVPTVTNVVLVLLDNVDNILDMIYFA
                     AKEIRRGRHLWFNDAIYSIVSAELKEMLDLVRTMLEYLQTMIKTNKAEDLRRYARIIS
                     SLEEEIDEIKENILDKVYALSLNAVEFNHIVSLIYTADKIADNIQDAAYHLATVVTSV
                     *
gene                 4131. .4895
                     /gene="PAE3188"
CDS                  4131. .4895
                     /gene="PAE3188"
                     /note="Hypothetical"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAL64737.1"
                     /db_xref="GI:18161455"
                     /translation="MELYGLILAALLIYNNSLVLLGPIAWGAGIGARKAFVIAAAAQL
                     LGVLTSTMAQLPISLPEFLYIGALYLLSLVKISLPISVIGYSIHAPRPEAVALWLLS
                     PLVSVATVAIQKTLKEGRDLAALSLFLVWVLFGFNNVALFTPDAALAAAVVAGTYFG
                     LGFSSWVIDIAALRRFTAAAVNLTVALGALAGILLSVPISPTLVAYSSILAASYSQGM
                     RIIRVEKFAAATLATLLALAAALIPYLKSLLAPRA"
gene                 complement(4852. .5523)
                     /gene="PAE3189"
CDS                  complement(4852. .5523)
                     /gene="PAE3189"
                     /note="Hypothetical; Conserved"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAL64738.1"
                     /db_xref="GI:18161456"
                     /translation="MPPSTGELLISWLGHDSPRITGGGPVLYIDPYQLQVGDPKADAI
                     LITHEHPDHCDPPSIQRILKPSTVVVAPRVARQCALKAANNVAIISPGEEREVGPLKI
                     RAFPAYNINKFRDPAGRVVFHPKEDGRVAYLIEWGGVRIPHAGDSDPVPEFREVRADV
                     VLVPVSGVTVMTPQEAAEFVNAVMPKVAIPMHYGSIVASRKEAEEFKRLVRPEVQVVI
                     LDREI"
gene                 complement(5555. .5728)
                     /gene="PAE3192"
CDS                  complement(5555. .5728)
                     /gene="PAE3192"
                     /note="Hypothetical; Conserved"
                     /codon_start=1
                     /transl_table=11
                     /product="paREP4"
                     /protein_id="AAL64739.1"
                     /db_xref="GI:18161457"
                     /translation="MSKKQKLKFYDIKAKQAFETDQYEVIEKQTARGPMMPAVAKSPY
                     TGIKVFRLLGKKK"

gene                 5812. .5901
                     /gene="PAE3193"
CDS                  5812. .5901
                     /gene="PAE3193"
                     /note="Hypothetical"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAL64740.1"
                     /db_xref="GI:18161458"
                     /translation="MDTGLLALLYIAALAASIHFVPRTAGARR"
gene                 5923. .6342
                     /gene="PAE3193a"
CDS                  5923. .6342
                     /gene="PAE3193a"
                     /note="Cellular processes; Adaptations to atypical
                     conditions"
                     /codon_start=1
                     /transl_table=11
                     /product="possible heat shock protein, htpX homolog part
                     1, authentic frameshift"
                     /protein_id="AAL64741.1"
                     /db_xref="GI:18161459"
                     /translation="MALILAAIVAVAVAVLAPLLPPAVVLTQAAAIGIKDYITAFAVLA
                     AVSAAAMYIISPYIINAVFGPRPDPGLQQLVDAVAAKLGGRVKARAVVVEGPPKRLRL
                     RQLPNGSVRGGDHGLVKHGEPRRAQGRHRPRAGTPHK"
gene                 6254. .6838
                     /gene="PAE3195"
CDS                  6254. .6838
                     /gene="PAE3195"
                     /note="Cellular processes; Adaptations to atypical
                     conditions"
                     /codon_start=1
                     /transl_table=11
                     /product="possible heat shock protein, htpX homolog part
                     2, authentic frameshift"
                     /protein_id="AAL64742.1"
                     /db_xref="GI:18161460"
                     /translation="MAVPTGLLNTASRDELEAVIGHELGHHINKDMPIMTAIGLLPSI
                     AFPTGVAAIHLGTLADRREPSILAIAYGVVMIVVSFIVQLLVLSFSKLRETYADMHGAR
                     AAGSDPQAMMNALAKIHKYTAQNPEAIATSPKISGFKALFIYALINAAANPLIDITPEEV
                     RQLMRLQTSWLDEILSSSHPPIPKRLKVLSALAAP"
gene                 6928. .7006
                     /gene="PAEtRl4"
tRNA                 6928. .7006
                     /gene="PAEtRl4"
                     /product="tRNA-Gly"
gene                 7139. .8140
                     /gene="PAE3196"
CDS                  7139. .8140
                     /gene="PAE3196"
                     /note="Hypothetical"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAL64743.1"
                     /db_xref="GI:18161461"
                     /translation="MGVEICRSLLWCLGALGRSQRLYAAAGLVDEEGLEAASRAAGEL
                     RVLVGDSGPVPRVTERNREVVEVYPSIHAKFYIPAEDAGPSAALVGSADLTAGGLRG
                     RLEAVVLRGSAARPLADMFNRLWARALPLTEDVVADWEPEEALEKGWGEAVKRANE
                     RLAETLGVSAHCLSRHDPLNCARLYARAVRSRFEGCGDLPENCAARATGVSAKALLSA
                     PPSAVLAGHYYCWARALAARLLEGKVGRLDSGMEAYEAAVQAGAESCWGEAKRAAEEE
                     LERLEDSNYRDNVVRWPIPTKLIFLAMTLPATGCRILGREVRTKKRGVARVEKELYC'
gene                 complement(8131. .10143)
                     /gene="PAE3197"
CDS                  complement(8131. .10143)

Query Match          82.0%;  Score 16.4;  DB 1;  Length 14086;
Best Local Similarity  94.4%;  Pred. No. 7.2e+02;
Matches  17;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy       1 ATATTAACGCCCACGCTC 18
           ||||||||||||||||||
```

```
Db    5813 ATATTAACGCCACGCCC 5796


RESULT 13
AC014130/c
LOCUS       AC014130               40123 bp    DNA     linear   HTG 16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC014130
VERSION     AC014130.1  GI:6437205
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 40123)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10210776 by the submitter.
            For further information on this sequence e-mail to flyc@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * this sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..40123
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    10427 a    9574 c    9411 g   10711 t
ORIGIN

  Query Match          82.0%;  Score 16.4;  DB 2;  Length 40123;
  Best Local Similarity  94.4%;  Pred. No. 7.1e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 ATTAACGCCCACGCTCTC 20
           |||||||||| ||||||||
Db   39472 ATTAACGCCCCCGCTCTC 39455


RESULT 14
AC016523
LOCUS       AC016523               65961 bp    DNA     linear   HTG 02-DEC-1999
DEFINITION  Drosophila melanogaster chromosome X clone BAC%40P05 (D120B)
            RPCI-98 40.P.5 map 16B-16B strain y; cn bw sp, *** SEQUENCING IN
            PROGRESS ***, 60 unordered pieces.
ACCESSION   AC016523
VERSION     AC016523.1  GI:6503262
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 65961)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Sutenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
            Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
            Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 65961)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Sutenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
            Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
            Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-DEC-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 200 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 60 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *      1      656: contig of 656 bp in length
            *    657      736: gap of unknown length
            *    737     1733: contig of 997 bp in length
            *   1734     1813: gap of unknown length
            *   1814     2418: contig of 605 bp in length
            *   2419     2498: gap of unknown length
            *   2499     3364: contig of 866 bp in length
            *   3365     3444: gap of unknown length
            *   3445     4009: contig of 565 bp in length
            *   4010     4089: gap of unknown length
            *   4090     4784: contig of 695 bp in length
            *   4785     4864: gap of unknown length
            *   4865     5717: contig of 853 bp in length
            *   5718     5797: gap of unknown length
            *   5798     6515: contig of 718 bp in length
            *   6516     6595: gap of unknown length
            *   6596     7319: contig of 724 bp in length
            *   7320     7399: gap of unknown length
            *   7400     8205: contig of 806 bp in length
            *   8206     8285: gap of unknown length
            *   8286     8954: contig of 669 bp in length
            *   8955     9034: gap of unknown length
            *   9035    10371: contig of 1337 bp in length
            *  10372    10451: gap of unknown length
            *  10452    11236: contig of 785 bp in length
            *  11237    11316: gap of unknown length
            *  11317    12306: contig of 990 bp in length
            *  12307    12386: gap of unknown length
            *  12387    13131: contig of 745 bp in length
            *  13132    13211: gap of unknown length
            *  13212    14351: contig of 1140 bp in length
            *  14352    14431: gap of unknown length
            *  14432    15050: contig of 619 bp in length
            *  15051    15130: gap of unknown length
            *  15131    16458: contig of 1328 bp in length
            *  16459    16538: gap of unknown length
            *  16539    17995: contig of 1457 bp in length
            *  17996    18075: gap of unknown length
            *  18076    19235: contig of 1160 bp in length
            *  19236    19315: gap of unknown length
            *  19316    21063: contig of 1748 bp in length
            *  21064    21143: gap of unknown length
            *  21144    22659: contig of 1516 bp in length
            *  22660    22739: gap of unknown length
            *  22740    24028: contig of 1289 bp in length
            *  24029    24108: gap of unknown length
            *  24109    24779: contig of 671 bp in length
            *  24780    24859: gap of unknown length
            *  24860    26341: contig of 1482 bp in length
            *  26342    26421: gap of unknown length
            *  26422    27687: contig of 1266 bp in length
            *  27688    27767: gap of unknown length
            *  27768    29294: contig of 1527 bp in length
```

```
     *    29295    29374: gap of unknown length
     *    29375    30634: contig of 1260 bp in length
     *    30635    30714: gap of unknown length
     *    30715    32904: contig of 2190 bp in length
     *    32905    32984: gap of unknown length
     *    32985    34623: contig of 1639 bp in length
     *    34624    34703: gap of unknown length
     *    34704    36780: contig of 2077 bp in length
     *    36781    36860: gap of unknown length
     *    36861    38926: contig of 2066 bp in length
     *    38927    39006: gap of unknown length
     *    39007    41830: contig of 2824 bp in length
     *    41831    41910: gap of unknown length
     *    41911    45034: contig of 3124 bp in length
     *    45035    45114: gap of unknown length
     *    45115    49237: contig of 4123 bp in length
     *    49238    49317: gap of unknown length
     *    49318    49969: contig of 652 bp in length
     *    49970    50049: gap of unknown length
     *    50050    50691: contig of 642 bp in length
     *    50692    50771: gap of unknown length
     *    50772    51393: contig of 622 bp in length
     *    51394    51473: gap of unknown length
     *    51474    52111: contig of 638 bp in length
     *    52112    52191: gap of unknown length
     *    52192    52836: contig of 645 bp in length
     *    52837    52916: gap of unknown length
     *    52917    53464: contig of 548 bp in length
     *    53465    53544: gap of unknown length
     *    53545    53758: contig of 214 bp in length
     *    53759    53838: gap of unknown length
     *    53839    54430: contig of 592 bp in length
     *    54431    54510: gap of unknown length
     *    54511    55156: contig of 646 bp in length
     *    55157    55236: gap of unknown length
     *    55237    55863: contig of 627 bp in length
     *    55864    55943: gap of unknown length
     *    55944    56516: contig of 573 bp in length
     *    56517    56596: gap of unknown length
     *    56597    57180: contig of 584 bp in length
     *    57181    57260: gap of unknown length
     *    57261    57872: contig of 612 bp in length
     *    57873    57952: gap of unknown length
     *    57953    58653: contig of 701 bp in length
     *    58654    58733: gap of unknown length
     *    58734    59364: contig of 631 bp in length
     *    59365    59444: gap of unknown length
     *    59445    60126: contig of 682 bp in length
     *    60127    60206: gap of unknown length
     *    60207    60835: contig of 629 bp in length
     *    60836    60915: gap of unknown length
     *    60916    61535: contig of 620 bp in length
     *    61536    61615: gap of unknown length
     *    61616    62021: contig of 406 bp in length
     *    62022    62101: gap of unknown length
     *    62102    62937: contig of 836 bp in length
     *    62938    63017: gap of unknown length
     *    63018    63659: contig of 642 bp in length
     *    63660    63739: gap of unknown length
     *    63740    64372: contig of 633 bp in length
     *    64373    64452: gap of unknown length
     *    64453    65031: contig of 579 bp in length
     *    65032    65111: gap of unknown length
     *    65112    65473: contig of 362 bp in length
     *    65474    65553: gap of unknown length
     *    65554    65961: contig of 408 bp in length.
FEATURES             Location/Qualifiers
     source          1..65961
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="X"
                     /map="16B-16B"
                     /clone="BACR40P05 (DI208) RPCI-98 40.P.5"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBAC3.6"
BASE COUNT    16392 a  14437 c  13716 g  16689 t   4727 others
ORIGIN

  Query Match           82.0%;  Score 18.4;  DB 2;  Length 65961;
  Best Local Similarity   94.4%;  Pred. No. 7e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 ATTAACGCCCACGCTCTC 20
          |||||||||| |||||||
Db  50448 ATTAACGCCCCCGCTCTC 50465


RESULT 15
AC012649/c
LOCUS       AC012649              80866 bp    DNA     linear   HTG 04-NOV-1999
DEFINITION  Drosophila melanogaster chromosome X clone BACR07N08 (DI121)
            RPCI-98 07.N.8 map 17D-17E strain y; cn bw sp, *** SEQUENCING IN
            PROGRESS ***, 101 unordered pieces.
ACCESSION   AC012649
VERSION     AC012649.2  GI:6226990
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 80866)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
            Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
            Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 80866)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hinkle,A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Shir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Nov 4, 1999 this sequence version replaced gi:6175114.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgpfruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 200 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 101 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1      577: contig of 577 bp in length
            *     578      657: gap of unknown length
            *     658     1404: contig of 747 bp in length
            *    1405     1484: gap of unknown length
            *    1485     2313: contig of 829 bp in length
```

```
*     2314     2393: gap of unknown length
*     2394     3131: contig of 738 bp in length
*     3132     3211: gap of unknown length
*     3212     3719: contig of 508 bp in length
*     3720     3799: gap of unknown length
*     3800     4897: contig of 1098 bp in length
*     4898     4977: gap of unknown length
*     4978     5799: contig of 822 bp in length
*     5800     5879: gap of unknown length
*     5880     6469: contig of 590 bp in length
*     6470     6549: gap of unknown length
*     6550     7442: contig of 893 bp in length
*     7443     7522: gap of unknown length
*     7523     8486: contig of 964 bp in length
*     8487     8566: gap of unknown length
*     8567     9246: contig of 680 bp in length
*     9247     9326: gap of unknown length
*     9327     9971: contig of 645 bp in length
*     9972    10051: gap of unknown length
*    10052    10632: contig of 581 bp in length
*    10633    10712: gap of unknown length
*    10713    11328: contig of 616 bp in length
*    11329    11408: gap of unknown length
*    11409    12070: contig of 662 bp in length
*    12071    12150: gap of unknown length
*    12151    12954: contig of 804 bp in length
*    12955    13034: gap of unknown length
*    13035    13887: contig of 853 bp in length
*    13888    13967: gap of unknown length
*    13968    14675: contig of 708 bp in length
*    14676    14755: gap of unknown length
*    14756    15777: contig of 1022 bp in length
*    15778    15857: gap of unknown length
*    15858    16561: contig of 704 bp in length
*    16562    16641: gap of unknown length
*    16642    17656: contig of 1015 bp in length
*    17657    17736: gap of unknown length
*    17737    18347: contig of 611 bp in length
*    18348    18427: gap of unknown length
*    18428    19269: contig of 842 bp in length
*    19270    19349: gap of unknown length
*    19350    20123: contig of 774 bp in length
*    20124    20203: gap of unknown length
*    20204    21037: contig of 834 bp in length
*    21038    21117: gap of unknown length
*    21118    21880: contig of 763 bp in length
*    21881    21960: gap of unknown length
*    21961    23063: contig of 1103 bp in length
*    23064    23143: gap of unknown length
*    23144    23742: contig of 599 bp in length
*    23743    23822: gap of unknown length
*    23823    24697: contig of 875 bp in length
*    24698    24777: gap of unknown length
*    24778    25699: contig of 922 bp in length
*    25700    25779: gap of unknown length
*    25780    26547: contig of 768 bp in length
*    26548    26627: gap of unknown length
*    26628    27682: contig of 1055 bp in length
*    27683    27762: gap of unknown length
*    27763    28171: contig of 409 bp in length
*    28172    28251: gap of unknown length
*    28252    28823: contig of 572 bp in length
*    28824    28903: gap of unknown length
*    28904    29965: contig of 1062 bp in length
*    29966    30045: gap of unknown length
*    30046    31173: contig of 1128 bp in length
*    31174    31253: gap of unknown length
*    31254    32156: contig of 903 bp in length
*    32157    32236: gap of unknown length
*    32237    32872: contig of 636 bp in length
*    32873    32952: gap of unknown length
*    32953    33817: contig of 865 bp in length
*    33818    33897: gap of unknown length
*    33898    35066: contig of 1169 bp in length
*    35067    35146: gap of unknown length
*    35147    37188: contig of 2042 bp in length
*    37189    37268: gap of unknown length
*    37269    38406: contig of 1138 bp in length
*    38407    38486: gap of unknown length
*    38487    40075: contig of 1589 bp in length
*    40076    40155: gap of unknown length
*    40156    41204: contig of 1049 bp in length
*    41205    41284: gap of unknown length
*    41285    42613: contig of 1329 bp in length
*    42614    42693: gap of unknown length
*    42694    45350: contig of 2657 bp in length
*    45351    45430: gap of unknown length
*    45431    45914: contig of 484 bp in length
*    45915    45994: gap of unknown length
*    45995    46589: contig of 595 bp in length
*    46590    46669: gap of unknown length
*    46670    47222: contig of 553 bp in length
*    47223    47302: gap of unknown length
*    47303    47887: contig of 585 bp in length
*    47888    47967: gap of unknown length
*    47968    48553: contig of 586 bp in length
*    48554    48633: gap of unknown length
*    48634    49266: contig of 633 bp in length
*    49267    49346: gap of unknown length
*    49347    49944: contig of 598 bp in length
*    49945    50024: gap of unknown length
*    50025    50642: contig of 618 bp in length
*    50643    50722: gap of unknown length
*    50723    51227: contig of 505 bp in length
*    51228    51307: gap of unknown length
*    51308    52034: contig of 727 bp in length
*    52035    52114: gap of unknown length
*    52115    52686: contig of 572 bp in length
*    52687    52766: gap of unknown length
*    52767    53363: contig of 597 bp in length
*    53364    53443: gap of unknown length
*    53444    53993: contig of 550 bp in length
*    53994    54073: gap of unknown length
*    54074    54434: contig of 361 bp in length
*    54435    54514: gap of unknown length
*    54515    55129: contig of 615 bp in length
*    55130    55209: gap of unknown length
*    55210    55887: contig of 678 bp in length
*    55888    55967: gap of unknown length
*    55968    56510: contig of 543 bp in length
*    56511    56590: gap of unknown length
*    56591    57022: contig of 432 bp in length
*    57023    57102: gap of unknown length
*    57103    57394: contig of 292 bp in length
*    57395    57474: gap of unknown length
*    57475    58106: contig of 632 bp in length
*    58107    58186: gap of unknown length
*    58187    58920: contig of 734 bp in length
*    58921    59000: gap of unknown length
*    59001    59609: contig of 609 bp in length
*    59610    59689: gap of unknown length
*    59690    60371: contig of 682 bp in length
*    60372    60451: gap of unknown length
*    60452    61001: contig of 550 bp in length
*    61002    61081: gap of unknown length
*    61082    61656: contig of 575 bp in length
*    61657    61736: gap of unknown length
*    61737    62261: contig of 525 bp in length
*    62262    62341: gap of unknown length
*    62342    62865: contig of 524 bp in length
*    62866    62945: gap of unknown length
*    62946    63538: contig of 593 bp in length
*    63539    63618: gap of unknown length
*    63619    64136: contig of 518 bp in length
*    64137    64216: gap of unknown length
```

Query Match 82.0%; Score 16.4; DB 2; Length 80666;
Best Local Similarity 94.4%; Pred. No. 7e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       3 ATTAACGCCCACGCTCTC 20
           |||||||||| |||||||
Db   62105 ATTAACGCCCCGCTCTC 62088
```

Search completed: May 12, 2003, 01:53:59
Job time : 244 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 00:45:59 ; Search time 35.2143 Seconds
(without alignments)
1279.025 Million cell updates/sec

Title: US-09-980-277-3
Perfect score: 20
Sequence: 1 atattaacgcccacgctctc 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	20	100.0	20	22	AAI70402	Schistosoma manson
c	2	20	100.0	120	22	AAI70400	Schistosoma manson
c	3	20	100.0	6494	21	AAA46170	GFP-Sm1-7 fusion p
c	4	16.8	84.0	747	24	ABQ22068	Oligonucleotide fo
	5	16.8	84.0	747	24	ABQ22069	Oligonucleotide fo
c	6	16.8	84.0	748	24	ABQ53094	Oligonucleotide fo
	7	16.8	84.0	748	24	ABQ53095	Oligonucleotide fo
c	8	16.4	82.0	3226	22	AAH18430	Human cDNA sequenc
c	9	16.4	82.0	3251	23	ABL13875	Drosophila melanog
	10	16.4	82.0	5484	23	ABL13874	Drosophila melanog
c	11	15.8	79.0	729	23	AAS66682	DNA encoding novel
c	12	15.8	79.0	729	23	AAS75819	DNA encoding novel
c	13	15.8	79.0	890	24	ABQ34482	Oligonucleotide fo
	14	15.8	79.0	890	24	ABQ34483	Oligonucleotide fo
	15	15.8	79.0	3759	23	ABL05751	Drosophila melanog
	16	15.8	79.0	6020	23	ABL05750	Drosophila melanog
c	17	15.8	79.0	6075	24	ABJ33551	Human immune syste
c	18	15.8	79.0	6207	24	AAS63341	Chemically pretrea
c	19	15.8	79.0	6922	22	AAS46533	Tumour suppressor
c	20	15.8	79.0	6922	24	ABL70379	Chemically treated
c	21	15.8	79.0	6922	24	AAS61331	Human gene regulat
c	22	15.8	79.0	6922	24	ABK31410	Signal transductio
c	23	15.8	79.0	7631	24	ABL32860	Human immune syste
c	24	15.8	79.0	8076	24	ABK39954	Human chemically p
c	25	15.8	79.0	9543	24	ABN80049	Human chemically m
c	26	15.8	79.0	19459	24	ABL70528	Chemically treated
c	27	15.8	79.0	19459	24	ABK31213	Signal transductio
c	28	15.8	79.0	33053	24	ABN07006	Human angiogenesis
c	29	15.8	79.0	34548	24	ABL70604	Chemically treated
c	30	15.4	77.0	336	18	AAT67489	H. pylori cytoplas
c	31	15.4	77.0	2535	23	AAS53763	Helicobacter pylor
c	32	15.4	77.0	2547	18	AAT67908	H. pylori cytoplas
c	33	15.4	77.0	2577	22	AAF25651	H. pylori HPN048 e
c	34	15.4	77.0	2577	23	AAS53662	Helicobacter pylor
c	35	15.4	77.0	2627	19	AAX14411	H. pylori GHPO 615
c	36	15.4	77.0	2627	19	AAV52087	Helicobacter polyp
c	37	15.2	76.0	103	23	ABV14698	Human prostate exp
c	38	15.2	76.0	105	23	ABV03444	Human prostate exp
c	39	15.2	76.0	124	22	AAL23967	Human breast cance
c	40	15.2	76.0	130	22	AAL15670	Human breast cance
c	41	15.2	76.0	139	23	ABV03603	Human prostate exp
c	42	15.2	76.0	144	23	ABV05529	Human prostate exp
c	43	15.2	76.0	149	23	ABV07191	Human prostate exp
c	44	15.2	76.0	165	23	ABV05266	Human prostate exp
	45	15.2	76.0	179	22	AAL19730	Human breast cance

ALIGNMENTS

RESULT 1
AAI70402
ID AAI70402 standard; DNA; 20 BP.
XX
AC AAI70402;
XX
DT 07-JAN-2002 (first entry)
XX
DE Schistosoma mansoni repeater unit 3' PCR primer.
XX
KW Schistosomiasis; infection; diagnosis; PCR primer; ss.
XX
OS Schistosoma mansoni.
XX
PN WO200175148-A1.
XX
PD 11-OCT-2001.
XX
PF 04-APR-2001; 2001WO-BR00035.
XX
PR 04-APR-2000; 2000BR-0001536.
XX
PA (FIOC-) FIOCRUZ FUNDACAO CRUZ OSWALDO.
XX
PI Teles Rabello AL, Dias Neto E, Pontes LA;
XX
DR WPI; 2001-648561/74.
XX
PT Diagnosing infection by parasites of the Schistosoma sp. in a
PT biological sample especially useful in cases of low infection
PT intensity, comprises detecting a specific region of the DNA of
PT Schistosoma by polymerase chain reaction -

```
XX
PS   Claim 4; Page 13; 34pp; English.
XX
CC   The present sequence is that of a claimed oligonucleotide primer,
CC   derived from the 3' region of a highly repeated short DNA sequence
CC   (see AAI70400) of the Schistosoma mansoni genome.  This 3' primer,
CC   and the 5' primer given in AAI70401, are used in the method of the
CC   invention for diagnosing Schistosoma infection.  The method involves
CC   collection of the sample to be examined, extraction of Schistosoma
CC   sp. DNA, amplification by PCR, separation of PCR products by
CC   electrophoresis, and detection e.g. by colouring with silver salts.
CC   A kit for diagnosing infection, which includes the primers, is
CC   claimed.  The method is useful for detecting Schistosoma sp.
CC   parasites by detecting the parasite's DNA in a biological sample.
CC   It is especially useful in cases of low infection intensity for
CC   which parasitological stool tests demonstrate little sensitivity.
XX
SQ   Sequence 20 BP; 5 A; 8 C; 2 G; 5 T; 0 other;

  Query Match             100.0%;  Score 20;  DB 22;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 0.38;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCTC 20
           ||||||||||||||||||||
Db       1 ATATTAACGCCCACGCTCTC 20


RESULT 2
AAI70400/c
ID   AAI70400 standard; DNA; 120 BP.
XX
AC   AAI70400;
XX
DT   07-JAN-2002  (first entry)
XX
DE   Schistosoma mansoni repeater unit DNA.
XX
KW   Schistosomiasis; infection; diagnosis; ds.
XX
OS   Schistosoma mansoni.
XX
FH   Key             Location/Qualifiers
FT   primer_bind     complement (1..19)
FT                   /*tag=  a
FT                   /note= "primer of AAI70401"
FT   primer_bind     91..110
FT                   /*tag=  b
FT                   /note= "primer of AAI70402"
XX
PN   WO200175148-A1.
XX
PD   11-OCT-2001.
XX
PF   04-APR-2001; 2001WO-BR00035.
XX
PR   04-APR-2000; 2000BR-0001536.
XX
PA   (FIOC-) FIOCRUZ FUNDACAO CRUZ OSWALDO.
XX
PI   Teles Rabello AL,  Dias Neto E,  Pontes LA;
XX
DR   WPI; 2001-648561/74.
XX
PT   Diagnosing infection by parasites of the Schistosoma sp. in a
PT   biological sample especially useful in cases of low infection
PT   intensity, comprises detecting a specific region of the DNA of
PT   Schistosoma by polymerase chain reaction -
XX
PS   Claim 1; Page 22; 34pp; English.
XX
CC   The present sequence is that of a highly repeated short DNA
CC   sequence in the genome of Schistosoma mansoni.  PCR primers
CC   (see AAI70401 and AAI70402) that flank this sequence are used in
CC   the method of the invention for diagnosing Schistosoma infection.
CC   The method involves collection of the sample to be examined,
CC   extraction of Schistosoma sp. DNA, amplification by PCR, separation
CC   of PCR products by electrophoresis, and detection e.g. by colouring
CC   with silver salts.  A kit for diagnosing infection, which includes
CC   the primers, is claimed.  The method is useful for detecting
CC   Schistosoma sp. parasites by detecting the DNA of the parasite in a
CC   biological sample.  It is especially useful in cases of low
CC   infection intensity for which parasitological stool tests
CC   demonstrate little sensitivity.
XX
SQ   Sequence 120 BP; 36 A; 20 C; 26 G; 38 T; 0 other;

  Query Match             100.0%;  Score 20;  DB 22;  Length 120;
  Best Local Similarity   100.0%;  Pred. No. 0.46;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCTC 20
           ||||||||||||||||||||
Db     110 ATATTAACGCCCACGCTCTC 91


RESULT 3
AAA46170/c
ID   AAA46170 standard; DNA; 6494 BP.
XX
AC   AAA46170;
XX
DT   27-SEP-2000  (first entry)
XX
DE   GFP-Sm1-7 fusion protein construct.
XX
KW   GFP-Sm1-7 fusion construct; circular; green fluorescent protein;
KW   glutathione S-transferase; eukaryotic diploid multicellular parasite;
KW   universal graft; transgenic eukaryotic parasite; acquired deficiency;
KW   genetic deficiency; hormone deficiency; metabolic deficiency;
KW   haematological deficiency; immunological deficiency; immunotherapy;
KW   anti-microbial therapy; anti-cancer therapy; drug addiction;
KW   poisoning condition; geriatric condition; ds.
XX
OS   Synthetic.
XX
PN   WO200032804-A1.
XX
PD   08-JUN-2000.
XX
PF   01-DEC-1999;  99WO-IL00651.
XX
PR   01-DEC-1998;  98US-0201850.
XX
PA   (YISS ) YISSUM RES & DEV CO.
XX
PI   Hamburger J,  Laban A;
XX
DR   WPI; 2000-412348/35.
XX
PT   Eukaryotic diploid multicellular parasite useful as universal grafts
PT   for in vivo delivery of beneficial gene products in humans and animals
PT   involves transformation with a transgene -
XX
PS   Example 2; Fig 6; 90pp; English.
XX
CC   This sequence represents a GFP-Sm1-7 fusion construct contained within a
CC   recombinant vector. This sequence contains the green fluorescent protein
CC   (GFP) coding sequence from Aequorea victoria, and the
CC   Sm1-7 coding sequence from Schistosoma, along with
CC   promoter sequences. The invention relates to a eukaryotic diploid
CC   multicellular parasite transformed with a transgene. Transgenic
CC   eukaryotic parasites are useful as universal grafts for in vivo delivery
CC   of beneficial gene product in humans and animals. The parasites can
CC   particularly be used for restoration of deficiencies whether acquired or
CC   genetic, such as hormone deficiencies, metabolic deficiencies,
```

```
CC   haematological deficiencies, immunological deficiencies, immunotherapy,
CC   anti-microbial therapy, anti-cancer therapy. They can also be used for
CC   treatment of drug addiction, of poisoning conditions, and for
CC   amelioration of geriatric conditions. Treatment of humans with in vivo
CC   transgenes are universally compatible, readily available and inexpensive.
CC   Genotypic alterations of the patients is avoided, reducing risks of
CC   mutagenesis and malignant transformation.
XX
SQ   Sequence 6494 BP; 1754 A; 1373 C; 1481 G; 1886 T; 0 other;

  Query Match             100.0%;  Score 20;  DB 21;  Length 6494;
  Best Local Similarity   100.0%;  Pred. No. 0.7;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCTC 20
           ||||||||||||||||||||
Db     110 ATATTAACGCCCACGCTCTC 91


RESULT 4
ABQ22068/c
ID   ABQ22068 standard; DNA; 747 BP.
XX
AC   ABQ22068;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 8659.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 747 BP; 99 A; 102 C; 297 G; 249 T; 0 other;

  Query Match             84.0%;  Score 16.8;  DB 24;  Length 747;
  Best Local Similarity   90.0%;  Pred. No. 28;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCTC 20
           ||||||||||||||  ||||
Db     400 ATATTAACGCCCACCGTCTC 381


RESULT 5
ABQ22069
ID   ABQ22069 standard; DNA; 747 BP.
XX
AC   ABQ22069;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 8660.
XX
KW   Human; Cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
```

```
CC   the disclosure of the invention.
XX
SQ   Sequence 747 BP; 249 A; 297 C; 102 G; 99 T; 0 other;

  Query Match             84.0%;  Score 16.8;  DB 24;  Length 747;
  Best Local Similarity   90.0%;  Pred. No. 28;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCTC 20
           ||||||||||||||  ||||
Db     348 ATATTAACGCCCACCGTCTC 367


RESULT 6
ABQ53094/C
ID   ABQ53094 standard; DNA; 748 BP.
XX
AC   ABQ53094;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 39685.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 748 BP; 104 A; 101 C; 282 G; 261 T; 0 other;

  Query Match             84.0%;  Score 16.8;  DB 24;  Length 748;
  Best Local Similarity   90.0%;  Pred. No. 28;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCTC 20
           ||||||||||||||  ||||
Db     310 ATATTAACGCCCACCGTCTC 291


RESULT 7
ABQ53095
ID   ABQ53095 standard; DNA; 748 BP.
XX
AC   ABQ53095;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 39686.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 748 BP; 261 A; 282 C; 101 G; 104 T; 0 other;

  Query Match             84.0%;  Score 16.8;  DB 24;  Length 748;
  Best Local Similarity   90.0%;  Pred. No. 28;
```

```
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCTC 20
           |||||||||||| ||||
Db     439 ATATTAACGCCCACCGTCTC 458


RESULT 8
AAH18430/c
ID   AAH18430 standard; cDNA; 3226 BP.
XX
AC   AAH18430;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:18513.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0218036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000;  2000JP-0118776.
PR   02-MAY-2000;  2000JP-0183767.
PR   09-JUN-2000;  2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 18513; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 3226 BP; 785 A; 894 C; 902 G; 645 T; 0 other;

  Query Match             82.0%;  Score 16.4;  DB 22;  Length 3226;
  Best Local Similarity   94.4%;  Pred. No. 53;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTC 18
           ||||| ||||||||||||||
Db     615 ATATTCACGCCCACGCTC 598


RESULT 9
ABL13875/c
ID   ABL13875 standard; cDNA; 3251 BP.
XX
AC   ABL13875;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 36107.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABD69772.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 36107; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16175-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3251 BP; 849 A; 820 C; 880 G; 702 T; 0 other;

  Query Match             82.0%;  Score 16.4;  DB 23;  Length 3251;
  Best Local Similarity   94.4%;  Pred. No. 53;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 ATTAACGCCCACGCTCTC 20
           |||||||||| |||||||
Db    2810 ATTAACGCCCCCGCTCTC 2793


RESULT 10
ABL13874
ID   ABL13874 standard; cDNA; 5484 BP.
XX
AC   ABL13874;
```

```
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 36104.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB69771.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 36104; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 5484 BP; 1402 A; 1280 C; 1322 G; 1480 T; 0 other;

  Query Match           82.0%;  Score 16.4;  DB 23;  Length 5484;
  Best Local Similarity  94.4%;  Pred. No. 56;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 ATTAACGCCCACGCTCTC 20
           |||||||||| |||||||
Db     652 ATTAACGCCCCGCTCTC 669


RESULT 11
AAS66682/c
ID   AAS66682 standard; cDNA; 729 BP.
XX
AC   AAS66682;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #2486.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PP   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG02495.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 2486; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 729 BP; 269 A; 175 C; 148 G; 137 T; 0 other;

  Query Match           79.0%;  Score 15.8;  DB 23;  Length 729;
  Best Local Similarity  89.5%;  Pred. No. 94;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCT 19
           ||||||| |||| ||||
Db     514 ATATTAACCCCCACTCTCT 496


RESULT 12
AAS75819/c
ID   AAS75819 standard; cDNA; 729 BP.
XX
AC   AAS75819;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #11623.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PP   30-MAR-2001; 2001WO-US08631.
```

```
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG11632.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity
XX
PS   Claim 1; SEQ ID No 11623; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 729 BP; 269 A; 175 C; 148 G; 137 T; 0 other;

  Query Match             79.0%;  Score 15.8;  DB 23;  Length 729;
  Best Local Similarity   89.5%;  Pred. No. 94;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCT 19
           |||||||| ||||| ||||
Db     514 ATATTAACCCCCACTCTCT 496


RESULT 13
ABQ34482/c
ID   ABQ34482 standard; DNA; 890 BP.
XX
AC   ABQ34482;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 21073.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 890 BP; 134 A; 106 C; 283 G; 367 T; 0 other;

  Query Match             79.0%;  Score 15.8;  DB 24;  Length 890;
  Best Local Similarity   89.5%;  Pred. No. 96;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 TATTAACGCCCACGCTCTC 20
           | |||||||| |||||||
Db      23 TCTTAACGCCCCCGCTCTC 5


RESULT 14
ABQ34483
ID   ABQ34483 standard; DNA; 890 BP.
XX
AC   ABQ34483;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 21074.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
```

```
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 890 BP; 367 A; 283 C; 106 G; 134 T; 0 other;

  Query Match             79.0%;  Score 15.8;  DB 24;  Length 890;
  Best Local Similarity   89.5%;  Pred. No. 96;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 TATTAACGCCCACGCTCTC 20
           | ||||||||| |||||||
Db     868 TCTTAACGCCCCCGCTCTC 886


RESULT 15
ABL05751
ID   ABL05751 standard; cDNA; 3759 BP.
XX
AC   ABL05751;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 11735.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB61648.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 11735; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3759 BP; 963 A; 927 C; 946 G; 923 T; 0 other;

  Query Match             79.0%;  Score 15.8;  DB 23;  Length 3759;
  Best Local Similarity   89.5%;  Pred. No. 1.1e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 TATTAACGCCCACGCTCTC 20
           | ||||||| |||||||| ||
Db    3044 TATTAACCCCCACGCTATC 3062


Search completed: May 12, 2003, 01:14:58
Job time : 37.2143 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 01:08:23 ; Search time 8.78571 Seconds
(without alignments)
698.126 Million cell updates/sec

Title: US-09-980-277-3
Perfect score: 20
Sequence: 1 atattaacgcccacgctctc 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	14.8	74.0	3636	3	US-09-074-579-2	Sequence 2, Appli
c	2	14.8	74.0	3636	4	US-09-388-774-2	Sequence 2, Appli
c	3	14.2	71.0	308	2	US-08-673-190A-4	Sequence 4, Appli
	4	14.2	71.0	1235	3	US-08-672-850-5	Sequence 5, Appli
	5	14.2	71.0	18318	1	US-08-414-926A-6	Sequence 6, Appli
	6	14.2	71.0	18318	2	US-08-926-922-6	Sequence 6, Appli
	7	14.2	71.0	18318	3	US-09-253-682-6	Sequence 6, Appli
	8	14.2	71.0	18318	4	US-09-527-657-6	Sequence 6, Appli
	9	13.8	69.0	1530	4	US-09-134-078-4	Sequence 4, Appli
c	10	13.8	69.0	1611	4	US-09-532-594B-14	Sequence 14, Appl
c	11	13.8	69.0	1872	4	US-09-532-594B-3	Sequence 3, Appli
c	12	13.8	69.0	1872	4	US-09-532-594B-15	Sequence 15, Appl
	13	13.8	69.0	2475	4	US-09-624-693A-20	Sequence 20, Appl
	14	13.8	69.0	3098	2	US-08-818-070-1	Sequence 1, Appli
	15	13.8	69.0	3098	2	US-08-723-585-1	Sequence 1, Appli
	16	13.8	69.0	4463	2	US-08-760-489-1	Sequence 1, Appli
	17	13.8	69.0	4463	2	US-08-760-489-3	Sequence 3, Appli
	18	13.8	69.0	4463	4	US-09-185-373-1	Sequence 1, Appli
	19	13.8	69.0	4463	4	US-09-185-373-3	Sequence 3, Appli
c	20	13.8	69.0	4767	4	US-09-532-594B-1	Sequence 1, Appli
c	21	13.8	69.0	50000	4	US-09-146-053-4	Sequence 4, Appli
	22	13.6	68.0	131	1	US-07-998-003A-92	Sequence 92, Appl
	23	13.6	68.0	131	1	US-08-453-274B-92	Sequence 92, Appl
	24	13.6	68.0	131	1	US-08-453-695A-92	Sequence 92, Appl
	25	13.6	68.0	131	1	US-08-268-161A-92	Sequence 92, Appl
	26	13.6	68.0	131	1	US-08-453-702A-92	Sequence 92, Appl
	27	13.6	68.0	131	4	US-09-099-639-92	Sequence 92, Appl
	28	13.6	68.0	131	5	PCT-US93-12588-92	Sequence 92, Appl
	29	13.6	68.0	131	5	PCT-US95-08071-92	Sequence 92, Appl
c	30	13.6	68.0	1117	4	US-08-858-207A-40	Sequence 40, Appl
	31	13.6	68.0	1503	4	US-09-172-952-6	Sequence 6, Appli
c	32	13.6	68.0	1509	2	US-08-481-337A-1	Sequence 1, Appli
c	33	13.6	68.0	1509	2	US-08-696-268B-1	Sequence 1, Appli
c	34	13.6	68.0	1509	5	PCT-US95-05467-1	Sequence 1, Appli
c	35	13.6	68.0	1596	5	PCT-US94-11328A-3	Sequence 3, Appli
c	36	13.6	68.0	1607	2	US-08-986-217-4	Sequence 4, Appli
	37	13.6	68.0	1858	4	US-09-521-780-4	Sequence 4, Appli
c	38	13.6	68.0	1984	4	US-09-382-256-1	Sequence 1, Appli
c	39	13.6	68.0	1984	4	US-09-395-115-1	Sequence 1, Appli
c	40	13.6	68.0	1984	4	US-08-436-265-1	Sequence 1, Appli
c	41	13.6	68.0	1984	4	US-09-679-187-1	Sequence 1, Appli
c	42	13.6	68.0	3285	4	US-09-206-942-68	Sequence 68, Appl
c	43	13.6	68.0	3663	4	US-09-206-942-27	Sequence 27, Appl
c	44	13.6	68.0	3681	4	US-09-206-942-25	Sequence 25, Appl
	45	13.6	68.0	3808	2	US-08-417-210A-79	Sequence 79, Appl

ALIGNMENTS

```
RESULT 1
US-09-074-579-2/c
; Sequence 2, Application US/09074579
; Patent No. 6001596
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Patterson, Chandra
;    TITLE OF INVENTION:  GROWTH-ASSOCIATED TRYPSIN-TYPE
;    TITLE OF INVENTION:  INHIBITOR HEAVY CHAIN PRECURSOR
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  Word Perfect 6.1/MS-DOS 6.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/074,579
;      FILING DATE:  HEREWITH
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cerrone, Michael C
;      REGISTRATION NUMBER:  39,132
;      REFERENCE/DOCKET NUMBER:  PF-0505 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3636 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  UTRSNOT02
;      CLONE:  688183
US-09-074-579-2

  Query Match                    74.0%;  Score 14.8;  DB 3;  Length 3636;
```

```
  Best Local Similarity   88.9%;  Pred. No. 57;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ATATTAACGCCCACGCTC 18
          ||||| ||| ||||||||
Db    660 ATATTCACGTCCACGCTC 643


RESULT 2
US-09-388-774-2/c
; Sequence 2, Application US/09388774
; Patent No. 6228991
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Patterson, Chandra
;    TITLE OF INVENTION:  GROWTH-ASSOCIATED TRYPSIN-TYPE
;    TITLE OF INVENTION:  INHIBITOR HEAVY CHAIN PRECURSOR
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  Word Perfect 6.1/MS-DOS 6.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/388,774
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/074,579
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cerrone, Michael C
;      REGISTRATION NUMBER:  39,132
;      REFERENCE/DOCKET NUMBER:  PF-0505 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3636 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  UTRSNOT02
;      CLONE:  688183
US-09-388-774-2

  Query Match             74.0%;  Score 14.8;  DB 4;  Length 3636;
  Best Local Similarity   88.9%;  Pred. No. 57;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ATATTAACGCCCACGCTC 18
          ||||| ||| ||||||||
Db    660 ATATTCACGTCCACGCTC 643


RESULT 3
US-08-673-190A-4/c
; Sequence 4, Application US/08673190A
; Patent No. 5985668
;  GENERAL INFORMATION:
;    APPLICANT:  Mattes, Ralf
;    APPLICANT:  Klein, Kathrin
;    APPLICANT:  Stegmaier, Sabine
;    TITLE OF INVENTION:  Sucrose Metabolism Mutants.
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/673,190A
;      FILING DATE:  27-JUN-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Forman, David S.
;      REGISTRATION NUMBER:  33,694
;      REFERENCE/DOCKET NUMBER:  06473.0001-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)408-4000
;      TELEFAX:  (202)408-4400
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  308 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Enterobacter species
US-08-673-190A-4

  Query Match             71.0%;  Score 14.2;  DB 2;  Length 308;
  Best Local Similarity   84.2%;  Pred. No. 94;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 TATTAACGCCCACGCTCTC 20
          ||| ||  |||||||||
Db    102 TATTGACCACCACGCTCTC 84


RESULT 4
US-08-672-850-5
; Sequence 5, Application US/08672850
; Patent No. 6140117
;  GENERAL INFORMATION:
;    APPLICANT:  Milbrandt, Jeffrey
;    APPLICANT:  Araki, Toshiyuki
;    TITLE OF INVENTION:  NINJURIN
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/672,850
;      FILING DATE:  24-JUL-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
```

```
;       REGISTRATION NUMBER:  38,304
;       REFERENCE/DOCKET NUMBER:  A-63610
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 781-1989
;       TELEFAX:  (415) 398-3249
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1235 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
US-08-672-850-5

  Query Match           71.0%;  Score 14.2;  DB 3;  Length 1235;
  Best Local Similarity   84.2%;  Pred. No. 1.1e+02;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCT 19
           | | ||| ||||||||||
Db    1192 ACAATAAAGCCCACGCTCT 1210


RESULT 5
US-08-414-926A-6
; Sequence 6, Application US/08414926A
; Patent No. 5721354
;  GENERAL INFORMATION:
;    APPLICANT:  Spaete, Richard
;    APPLICANT:  Cha, Tai-An
;    TITLE OF INVENTION:  NOVEL HUMAN CYTOMEGALOVIRUS
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooley Godward Castro Huddleson & Tatum
;      STREET:  5 Palo Alto Square
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306-2155
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/414,926A
;      FILING DATE:  March 31, 1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  AVIR-011/OOUS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-494-7622
;      TELEFAX:  415-857-0663
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  18318 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Human CMV
;      STRAIN:  Toledo
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  511..1281
;      OTHER INFORMATION:  /product = "UL133"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1401..2384
;      OTHER INFORMATION:  /product = "UL135"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2478..3197
;      OTHER INFORMATION:  /product = "UL136"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3283..3789
;      OTHER INFORMATION:  /product = "UL138"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  4355..4759
;      OTHER INFORMATION:  /product = "UL139"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  4944..5285
;      OTHER INFORMATION:  /product = "UL140"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  5558..6832
;      OTHER INFORMATION:  /product = "UL141"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  6908..7825
;      OTHER INFORMATION:  /product = "UL142"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  7813..8088
;      OTHER INFORMATION:  /product = "UL143"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  8468..8995
;      OTHER INFORMATION:  /product = "UL144"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  9327..9626
;      OTHER INFORMATION:  /product = "UL145"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  9910..10260
;      OTHER INFORMATION:  /product = "UL146"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  10328..10804
;      OTHER INFORMATION:  /product = "UL147"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  11106..12053
;      OTHER INFORMATION:  /product = "UL148"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  12133..12942
;      OTHER INFORMATION:  /product = "UL132"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  13569..14210
;      OTHER INFORMATION:  /product = "UL130"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  16216..16581
;      OTHER INFORMATION:  /product = "UL149"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1004..1528
;      OTHER INFORMATION:  /product = "UL134"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3063..3350
;      OTHER INFORMATION:  /product = "UL137"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  16337..18262
```

```
;      OTHER INFORMATION:  /product = "UL150"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  17752..18759
;      OTHER INFORMATION:  /product = "UL151"
US-08-414-926A-6

  Query Match             71.0%;  Score 14.2;  DB 1;  Length 18318;
  Best Local Similarity   84.2%;  Pred. No. 1.4e+02;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCT 19
           || |||||||||| || |||
Db   15055 ATGTTAACGCCCGCGGTCT 15073


RESULT 6
US-08-926-922-6
; Sequence 6, Application US/08926922
; Patent No. 5925751
;  GENERAL INFORMATION:
;    APPLICANT:  Spaete, Richard
;    APPLICANT:  Cha, Tai-An
;    TITLE OF INVENTION:  NOVEL HUMAN CYTOMEGALOVIRUS
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Luann Cserr Attorney at Law
;      STREET:  750 Arimo Avenue
;      CITY:  Oakland
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94610
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/926,922
;      FILING DATE:  September 10, 1997
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  AVIR 11A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  510-834-1448
;      TELEFAX:  510-839-7810
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  18318 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Human CMV
;      STRAIN:  Toledo
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  511..1281
;      OTHER INFORMATION:  /product = "UL133"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1401..2384
;      OTHER INFORMATION:  /product = "UL135"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2478..3157
;      OTHER INFORMATION:  /product = "UL136"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3283..3789
;      OTHER INFORMATION:  /product = "UL138"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  4355..4759
;      OTHER INFORMATION:  /product = "UL139"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  4944..5285
;      OTHER INFORMATION:  /product = "UL140"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  5558..6832
;      OTHER INFORMATION:  /product = "UL141"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  6908..7825
;      OTHER INFORMATION:  /product = "UL142"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  7813..8088
;      OTHER INFORMATION:  /product = "UL143"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  8468..8995
;      OTHER INFORMATION:  /product = "UL144"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  9327..9626
;      OTHER INFORMATION:  /product = "UL145"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  9910..10260
;      OTHER INFORMATION:  /product = "UL146"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  10328..10804
;      OTHER INFORMATION:  /product = "UL147"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  11106..12053
;      OTHER INFORMATION:  /product = "UL148"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  12133..12942
;      OTHER INFORMATION:  /product = "UL132"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  13569..14210
;      OTHER INFORMATION:  /product = "UL130"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  16216..16581
;      OTHER INFORMATION:  /product = "UL149"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1004..1528
;      OTHER INFORMATION:  /product = "UL134"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3063..3350
;      OTHER INFORMATION:  /product = "UL137"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  16337..18262
;      OTHER INFORMATION:  /product = "UL150"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  17752..18759
;      OTHER INFORMATION:  /product = "UL151"
US-08-926-922-6
```

```
  Query Match           71.0%;  Score 14.2;  DB 2;  Length 18318;
  Best Local Similarity   84.2%;  Pred. No. 1.4e+02;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCT 19
           || ||||||||| || |||
Db   15055 ATGTTAACGCCCGCGGTCT 15073


RESULT 7
US-09-253-682-6
; Sequence 6, Application US/09253682
; Patent No. 6040170
;  GENERAL INFORMATION:
;    APPLICANT:  Spaete, Richard
;    APPLICANT:  Cha, Tai-An
;    TITLE OF INVENTION:  NOVEL HUMAN CYTOMEGALOVIRUS
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Luann Cserr Attorney at Law
;      STREET:  750 Arimo Avenue
;      CITY:  Oakland
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94610
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/253,682
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/926,922
;      FILING DATE:  September 10, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  AVIR 11A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  510-834-1448
;      TELEFAX:  510-839-7810
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  18318 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Human CMV
;      STRAIN:  Toledo
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  511..1281
;      OTHER INFORMATION:  /product = "UL133"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1401..2384
;      OTHER INFORMATION:  /product = "UL135"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2478..3197
;      OTHER INFORMATION:  /product = "UL136"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3283..3789
;      OTHER INFORMATION:  /product = "UL138"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  4355..4759
;      OTHER INFORMATION:  /product = "UL139"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  4944..5285
;      OTHER INFORMATION:  /product = "UL140"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  5558..6832
;      OTHER INFORMATION:  /product = "UL141"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  6908..7825
;      OTHER INFORMATION:  /product = "UL142"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  7813..8088
;      OTHER INFORMATION:  /product = "UL143"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  8468..8995
;      OTHER INFORMATION:  /product = "UL144"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  9327..9626
;      OTHER INFORMATION:  /product = "UL145"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  9910..10260
;      OTHER INFORMATION:  /product = "UL146"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  10328..10804
;      OTHER INFORMATION:  /product = "UL147"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  11106..12053
;      OTHER INFORMATION:  /product = "UL148"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  12133..12942
;      OTHER INFORMATION:  /product = "UL132"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  13569..14210
;      OTHER INFORMATION:  /product = "UL130"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  16216..16581
;      OTHER INFORMATION:  /product = "UL149"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1004..1528
;      OTHER INFORMATION:  /product = "UL134"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3063..3350
;      OTHER INFORMATION:  /product = "UL137"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  16337..18262
;      OTHER INFORMATION:  /product = "UL150"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  17752..18759
;      OTHER INFORMATION:  /product = "UL151"
US-09-253-682-6

  Query Match           71.0%;  Score 14.2;  DB 3;  Length 18318;
  Best Local Similarity   84.2%;  Pred. No. 1.4e+02;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;
```

```
Qy      1 ATATTAACGCCCACGCTCT 19
          || |||||||||| || |||
Db  15055 ATGTTAACGCCCGCGGTCT 15073


RESULT 8
US-09-527-657-6
; Sequence 6, Application US/09527657
; Patent No. 6291236
;  GENERAL INFORMATION:
;       APPLICANT: Spaete, Richard
;                  Cha, Tai-An
;       TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
;       NUMBER OF SEQUENCES: 27
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Lusnn Cserr Attorney at Law
;            STREET: 750 Arimo Avenue
;            CITY: Oakland
;            STATE: CA
;            COUNTRY: USA
;            ZIP: 94610
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: Patentin Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/527,657
;            FILING DATE: 17-Mar-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/926,922
;            FILING DATE: September 10, 1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Cserr, Luann
;            REGISTRATION NUMBER: 31,822
;            REFERENCE/DOCKET NUMBER: AVIR 11A
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 510-834-1448
;            TELEFAX: 510-839-7810
;    INFORMATION FOR SEQ ID NO: 6:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 18318 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       HYPOTHETICAL: NO
;       ANTI-SENSE: NO
;       ORIGINAL SOURCE:
;            ORGANISM: Human CMV
;            STRAIN: Toledo
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  511..1281
;            OTHER INFORMATION:  /product = "UL133"
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  1401..2384
;            OTHER INFORMATION:  /product = "UL135"
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  2479..3197
;            OTHER INFORMATION:  /product = "UL136"
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  3283..3789
;            OTHER INFORMATION:  /product = "UL138"
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  4355..4759
;            OTHER INFORMATION:  /product = "UL139"
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  4944..5285
;            OTHER INFORMATION:  /product = "UL140"
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  5558..6832
;            OTHER INFORMATION:  /product = "UL141"
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  6908..7825
;            OTHER INFORMATION:  /product = "UL142"
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  7813..8088
;            OTHER INFORMATION:  /product = "UL143"
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  8468..8995
;            OTHER INFORMATION:  /product = "UL144"
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  9327..9626
;            OTHER INFORMATION:  /product = "UL145"
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  9910..10260
;            OTHER INFORMATION:  /product = "UL146"
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  10328..10804
;            OTHER INFORMATION:  /product = "UL147"
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  11106..12053
;            OTHER INFORMATION:  /product = "UL148"
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  12133..12942
;            OTHER INFORMATION:  /product = "UL132"
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  13569..14210
;            OTHER INFORMATION:  /product = "UL130"
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  16216..16581
;            OTHER INFORMATION:  /product = "UL149"
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  1004..1528
;            OTHER INFORMATION:  /product = "UL134"
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  3063..3350
;            OTHER INFORMATION:  /product = "UL137"
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  16337..18262
;            OTHER INFORMATION:  /product = "UL150"
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  17752..18759
;            OTHER INFORMATION:  /product = "UL151"
;    SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-527-657-6

  Query Match             71.0%;  Score 14.2;  DB 4;  Length 18318;
  Best Local Similarity   84.2%;  Pred. No. 1.4e+02;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATATTAACGCCCACGCTCT 19
          || |||||||||| || |||
Db  15055 ATGTTAACGCCCGCGGTCT 15073
```

```
RESULT 9
US-09-134-078-4
; Sequence 4, Application US/09134078
; Patent No. 6368844
;  GENERAL INFORMATION:
;    APPLICANT:  Eylina, Edward J.
;    TITLE OF INVENTION:  GLYCOSIDASE ENZYMES
;    NUMBER OF SEQUENCES:  72
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Gray Cary Ware & Freidenrich LLP
;      STREET:  4365 Executive Drive, Suite 1600
;      CITY:  San Diego
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92121
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/134,078
;      FILING DATE:  13-AUG-1998
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/949,026
;      FILING DATE:  10-OCT-1997
;      APPLICATION NUMBER:  60/056,916
;      FILING DATE:  06-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haile, Lisa A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  09010/024002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  858/677-1456
;      TELEFAX:  858/677-1465
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1530 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  1...1527
US-09-134-078-4

  Query Match             69.0%;  Score 13.8;  DB 4;  Length 1530;
  Best Local Similarity   88.2%;  Pred. No. 1.8e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 ATTAACGCCCACGCTCT 19
           || ||||||||||| ||
Db     727 ATAAACGCCCACGCACI 743


RESULT 10
US-09-532-594B-14/c
; Sequence 14, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT:  Chorini, John A.
; APPLICANT:  Kotin, Robert M.
; APPLICANT:  Safer, Brian
; APPLICANT:  Davidson, Beverly
; TITLE OF INVENTION:  AAV4 VECTOR AND USES THEREOF
; FILE REFERENCE:  14014.0252U2
; CURRENT APPLICATION NUMBER:  US/09/532,594B
; CURRENT FILING DATE:  2000-03-22
; NUMBER OF SEQ ID NOS:  22
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 14
;   LENGTH:  1611
;   TYPE:  DNA
;   ORGANISM:  Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION:  Description of Artificial Sequence;  No. 6468524e =
;   OTHER INFORMATION:  synthetic construct
;   NAME/KEY:  misc_feature
;   OTHER INFORMATION:  AAV4 Rep 68 gene
US-09-532-594B-14

  Query Match             69.0%;  Score 13.8;  DB 4;  Length 1611;
  Best Local Similarity   88.2%;  Pred. No. 1.8e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 TTAACGCCCACGCTCTC 20
           || ||||||||| ||||
Db     302 TTGACGCCCACGGTCTC 286


RESULT 11
US-09-532-594B-3/c
; Sequence 3, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT:  Chorini, John A.
; APPLICANT:  Kotin, Robert M.
; APPLICANT:  Safer, Brian
; APPLICANT:  Davidson, Beverly
; TITLE OF INVENTION:  AAV4 VECTOR AND USES THEREOF
; FILE REFERENCE:  14014.0252U2
; CURRENT APPLICATION NUMBER:  US/09/532,594B
; CURRENT FILING DATE:  2000-03-22
; NUMBER OF SEQ ID NOS:  22
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH:  1872
;   TYPE:  DNA
;   ORGANISM:  Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION:  Description of Artificial Sequence;  No. 6468524e =
;   OTHER INFORMATION:  synthetic construct
;   NAME/KEY:  CDS
;   LOCATION:  (1)...(1872)
;   NAME/KEY:  misc_feature
;   OTHER INFORMATION:  AAV4 Rep gene (full length)
US-09-532-594B-3

  Query Match             69.0%;  Score 13.8;  DB 4;  Length 1872;
  Best Local Similarity   88.2%;  Pred. No. 1.8e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 TTAACGCCCACGCTCTC 20
           || ||||||||| ||||
Db     302 TTGACGCCCACGGTCTC 286


RESULT 12
US-09-532-594B-15/c
; Sequence 15, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
;  APPLICANT:  Chorini, John A.
;  APPLICANT:  Kotin, Robert M.
;  APPLICANT:  Safer, Brian
;  APPLICANT:  Davidson, Beverly
;  TITLE OF INVENTION:  AAV4 VECTOR AND USES THEREOF
;  FILE REFERENCE:  14014.0252U2
;  CURRENT APPLICATION NUMBER:  US/09/532,594B
;  CURRENT FILING DATE:  2000-03-22
;  NUMBER OF SEQ ID NOS:  22
;  SOFTWARE:  FastSEQ for Windows Version 4.0
```

```
; SEQ ID NO 15
;   LENGTH: 1872
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e -
;   OTHER INFORMATION: synthetic construct
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: AAV4 Rep 78 gene
US-09-532-594B-15

  Query Match             69.0%;  Score 13.8;  DB 4;  Length 1872;
  Best Local Similarity   88.2%;  Pred. No. 1.8e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 TTAACGCCCACGCTCTC 20
           || ||||||||| ||||
Db     302 TTGACGCCCACGGTCTC 286


RESULT 13
US-09-624-693A-20
; Sequence 20, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
;  APPLICANT: Yoshida, Roberta
;  APPLICANT:  Kootstra, Anna
;  TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
;  TITLE OF INVENTION:  Polynucleotide Sequences and Methods of Obtaining and
;  TITLE OF INVENTION:  Using Same
;  FILE REFERENCE: 29479/00NSC
;  CURRENT APPLICATION NUMBER: US/09/624,693A
;  CURRENT FILING DATE:  2000-07-24
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
;   LENGTH: 2475
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: misc_difference
;   LOCATION: (13, 34, 46, 49, 51, 57, 59, 68, 69, 73, 75 - 77, 79, 82, 84,
;   OTHER INFORMATION: n = A or C or G or T; "n" indicates no consensus at that position
;   OTHER INFORMATION: Description of Artificial Sequence: Consensus
;   OTHER INFORMATION: Sequence of SEQ ID NOs: 12, 16, and 18
; Patent No. 6355468
US-09-624-693A-20

  Query Match             69.0%;  Score 13.8;  DB 4;  Length 2475;
  Best Local Similarity   83.3%;  Pred. No. 1.9e+02;
  Matches   15;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 ATTACGCCCACGCTCTC 20
           ||| |||||||||||  ||
Db    1162 ATTCACGCCCACGCNGTC 1179


RESULT 14
US-08-818-070-1
; Sequence 1, Application US/08818070
; Patent No. 5831019
;  GENERAL INFORMATION:
;    APPLICANT:  Tully, Timothy
;    APPLICANT:  Bolwig, Gert M.
;    TITLE OF INVENTION:  ASSOCIATIVE LEARNING AND THE linotte
;    TITLE OF INVENTION:  GENE
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAMILTON, BROOK, SMITH & REYNOLDS, PC
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/818,070
;      FILING DATE:  14-MAR-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/723,585
;      FILING DATE:  01-OCT-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  CSHL95-03pA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3098 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  89..2833
US-08-818-070-1

  Query Match             69.0%;  Score 13.8;  DB 2;  Length 3098;
  Best Local Similarity   88.2%;  Pred. No. 1.9e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 TTAACGCCCACGCTCTC 20
           |||| ||||||||||
Db     803 TTAAGCCCCACGCTCTC 819


RESULT 15
US-08-723-585-1
; Sequence 1, Application US/08723585
; Patent No. 5831057
;  GENERAL INFORMATION:
;    APPLICANT:  Tully, Timothy
;    APPLICANT:  Bolwig, Gert M.
;    TITLE OF INVENTION:  ASSOCIATIVE LEARNING AND THE linotte
;    TITLE OF INVENTION:  GENE
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAMILTON, BROOK, SMITH & REYNOLDS, PC
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/723,585
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  CSHL95-03pA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
```

```
;       TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3098 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  89..2833
US-08-723-585-1

  Query Match             69.0%;  Score 13.8;  DB 2;  Length 3098;
  Best Local Similarity   88.2%;  Pred. No. 1.9e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 TTAACGCCCACGCTCTC 20
           |||| |||||||||||
Db     803 TTAAGCCCCACGCTCTC 819


Search completed:  May 12, 2003, 02:38:44
Job time :  21.7857 secs
```

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 02:35:34 ; Search time 18.7143 Seconds
(without alignments)
1328.082 Million cell updates/sec

Title: US-09-980-277-3
Perfect score: 20
Sequence: 1 atattaacgcccacgctctc 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	15.4	77.0	2535	10	US-09-815-242-7400	Sequence 7400, Ap
c 2	15.4	77.0	2577	10	US-09-815-242-7299	Sequence 7299, Ap
3	15.2	76.0	382	10	US-09-978-574-857	Sequence 857, App
c 4	15.2	76.0	395	9	US-09-918-995-37010	Sequence 37010, A
c 5	15.2	76.0	1529	10	US-09-820-893-20	Sequence 20, Appl
c 6	15.2	76.0	1796	10	US-09-820-893-43	Sequence 43, Appl
c 7	15.2	76.0	2033	9	US-09-981-876-64	Sequence 64, Appl
c 8	15.2	76.0	2033	9	US-09-148-545-64	Sequence 64, Appl
c 9	15.2	76.0	2136	9	US-09-981-876-120	Sequence 120, App
c 10	15.2	76.0	2136	9	US-09-148-545-120	Sequence 120, App
c 11	15.2	76.0	2136	10	US-09-820-893-44	Sequence 44, Appl
c 12	15.2	76.0	2196	10	US-09-917-800A-1469	Sequence 1469, Ap
c 13	15.2	76.0	6061	9	US-10-239-676-40	Sequence 40, Appl
14	14.8	74.0	873	9	US-09-738-626-1562	Sequence 1562, Ap
c 15	14.8	74.0	2331	9	US-09-944-413-54	Sequence 54, Appl
c 16	14.8	74.0	2331	9	US-09-944-403-54	Sequence 54, Appl
c 17	14.8	74.0	2331	9	US-09-944-896-54	Sequence 54, Appl
c 18	14.8	74.0	2331	9	US-09-944-944-54	Sequence 54, Appl
c 19	14.8	74.0	2331	9	US-09-944-907-54	Sequence 54, Appl
c 20	14.8	74.0	2331	9	US-09-944-929-54	Sequence 54, Appl
c 21	14.8	74.0	2331	9	US-10-174-590-59	Sequence 59, Appl
c 22	14.8	74.0	2331	9	US-10-176-758-59	Sequence 59, Appl
c 23	14.8	74.0	2331	9	US-10-175-737-59	Sequence 59, Appl
c 24	14.8	74.0	2331	9	US-10-173-706-59	Sequence 59, Appl
c 25	14.8	74.0	2331	9	US-10-175-738-59	Sequence 59, Appl
c 26	14.8	74.0	2331	9	US-10-175-752-59	Sequence 59, Appl
c 27	14.8	74.0	2331	9	US-10-176-482-59	Sequence 59, Appl
c 28	14.8	74.0	2331	9	US-10-176-757-59	Sequence 59, Appl
c 29	14.8	74.0	2331	9	US-10-176-913-59	Sequence 59, Appl
c 30	14.8	74.0	2331	9	US-10-180-552-59	Sequence 59, Appl
c 31	14.8	74.0	2331	9	US-10-180-557-59	Sequence 59, Appl
c 32	14.8	74.0	2331	9	US-10-173-700-59	Sequence 59, Appl
c 33	14.8	74.0	2331	9	US-10-174-572-59	Sequence 59, Appl
c 34	14.8	74.0	2331	9	US-10-174-579-59	Sequence 59, Appl
c 35	14.8	74.0	2331	9	US-10-174-582-59	Sequence 59, Appl
c 36	14.8	74.0	2331	9	US-10-174-588-59	Sequence 59, Appl
c 37	14.8	74.0	2331	9	US-10-175-739-59	Sequence 59, Appl
c 38	14.8	74.0	2331	9	US-10-175-740-59	Sequence 59, Appl
c 39	14.8	74.0	2331	9	US-10-175-743-59	Sequence 59, Appl
c 40	14.8	74.0	2331	9	US-10-176-488-59	Sequence 59, Appl
c 41	14.8	74.0	2331	9	US-10-176-492-59	Sequence 59, Appl
c 42	14.8	74.0	2331	9	US-10-176-747-59	Sequence 59, Appl
c 43	14.8	74.0	2331	9	US-10-176-750-59	Sequence 59, Appl
c 44	14.8	74.0	2331	9	US-10-176-985-59	Sequence 59, Appl
c 45	14.8	74.0	2331	9	US-10-176-987-59	Sequence 59, Appl

ALIGNMENTS

RESULT 1
US-09-815-242-7400/c
; Sequence 7400, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: PatSEQ for Windows Version 4.0
; SEQ ID NO 7400
; LENGTH: 2535
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2535)
US-09-815-242-7400

```
  Query Match             77.0%;  Score 15.4;  DB 10;  Length 2535;
  Best Local Similarity   94.1%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 ATTAACGCCCACGCTCT 19
           ||| ||||||||||||||
Db    1638 ATTCACGCCCACGCTCT 1622


RESULT 2
US-09-815-242-7299/c
; Sequence 7299, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7299
;   LENGTH: 2577
;   TYPE: DNA
;   ORGANISM: Helicobacter pylori
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(2577)
US-09-815-242-7299

  Query Match             77.0%;  Score 15.4;  DB 10;  Length 2577;
  Best Local Similarity   94.1%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 ATTAACGCCCACGCTCT 19
           ||| ||||||||||||||
Db    1680 ATTCACGCCCACGCTCT 1664


RESULT 3
US-09-878-574-857
; Sequence 857, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
;  APPLICANT: Byrum, Joseph R.
;  APPLICANT:  La Rosa, Thomas J.
;  APPLICANT:  Thompson, Michael D.
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(15401)B
;  CURRENT APPLICATION NUMBER: US/09/878,574
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: 09/333,535
;  PRIOR FILING DATE: 1999-06-14
;  NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 857
;   LENGTH: 382
;   TYPE: DNA
;   ORGANISM: Glycine max
;   OTHER INFORMATION: Clone ID: LIB3028-045-Q1-B1-E8
US-09-878-574-857

  Query Match             76.0%;  Score 15.2;  DB 10;  Length 382;
  Best Local Similarity   85.0%;  Pred. No. 1.1e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCTC 20
           ||||| |||||| | |||||
Db      58 ATATTCACGCCCCCTCTCTC 77


RESULT 4
US-09-918-995-37010/c
; Sequence 37010, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 37010
;   LENGTH: 395
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-918-995-37010

  Query Match             76.0%;  Score 15.2;  DB 9;  Length 395;
  Best Local Similarity   85.0%;  Pred. No. 1.1e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCTC 20
           ||||||||| ||| | ||||
Db      30 ATATTAACGTCCATGTTCTC 11


RESULT 5
US-09-820-893-20/c
; Sequence 20, Application US/09820893
; Patent No. US20020076705A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 31 Human Secreted Proteins
;  FILE REFERENCE: P2033P1
;  CURRENT APPLICATION NUMBER: US/09/820,893
;  CURRENT FILING DATE:  2001-03-30
;  PRIOR APPLICATION NUMBER: 09/531,119
;  PRIOR FILING DATE: 2000-03-20
;  PRIOR APPLICATION NUMBER: 60/102,895
;  PRIOR FILING DATE: 1998-10-02
;  NUMBER OF SEQ ID NOS: 140
;  SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 20
;   LENGTH: 1529
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
```

```
;  LOCATION: (1505)
;  OTHER INFORMATION: n equals a,t,g, or c
US-09-820-893-20

  Query Match             76.0%;  Score 15.2;  DB 10;  Length 1529;
  Best Local Similarity   85.0%;  Pred. No. 1.2e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATATTAACGCCCACGCTCTC 20
          |||||||| ||| | ||||
Db    144 ATATTAACGTCCATGTCTC 125


RESULT 6
US-09-820-893-43/c
; Sequence 43, Application US/09820893
; Patent No. US20020076705A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 31 Human Secreted Proteins
;  FILE REFERENCE: PZ033P1
;  CURRENT APPLICATION NUMBER: US/09/820,893
;  CURRENT FILING DATE:  2001-03-30
;  PRIOR APPLICATION NUMBER: 09/531,119
;  PRIOR FILING DATE: 2000-03-20
;  PRIOR APPLICATION NUMBER: 60/102,895
;  PRIOR FILING DATE: 1998-10-02
;  NUMBER OF SEQ ID NOS: 140
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
;  LENGTH: 1796
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: SITE
;  LOCATION: (417)
;  OTHER INFORMATION: n equals a,t,g, or c
US-09-820-893-43

  Query Match             76.0%;  Score 15.2;  DB 10;  Length 1796;
  Best Local Similarity   85.0%;  Pred. No. 1.3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATATTAACGCCCACGCTCTC 20
          |||||||| ||| | ||||
Db     66 ATATTAACGTCCATGTCTC 47


RESULT 7
US-09-981-876-64/c
; Sequence 64, Application US/09981876
; Patent No. US20020164669A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 70 Human Secreted Proteins
;  FILE REFERENCE: PZ001P1
;  CURRENT APPLICATION NUMBER: US/09/981,876
;  CURRENT FILING DATE:  2001-10-19
;  PRIOR APPLICATION NUMBER: 09/148,545
;  PRIOR FILING DATE: 1998-09-04
;  PRIOR APPLICATION NUMBER: 60/040,162
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: 60/040,333
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: 60/038,621
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: 60/040,161
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: 60/040,626
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: 60/040,334
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: 60/040,336
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: 60/040,163
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: 60/047,615
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,600
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,597
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,502
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,633
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,583
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,617
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,618
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,503
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,592
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,581
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,584
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,500
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,587
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,492
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,598
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,613
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,582
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,596
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,612
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,632
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,601
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/043,580
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: 60/043,568
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: 60/043,314
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: 60/043,569
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: 60/043,311
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: 60/043,671
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: 60/043,674
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: 60/043,669
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: 60/043,312
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: 60/043,313
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: 60/043,672
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: 60/043,315
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: 60/048,974
;  PRIOR FILING DATE: 1997-06-06
```

; PRIOR APPLICATION NUMBER: 60/056,886
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,877
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,889
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,893
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,630
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,878
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,662
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,872
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,882
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,637
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,903
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,888
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,879
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,880
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,894
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,911
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,636
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,874
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,910
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,864
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,631
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,845
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,892
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/047,595
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/057,761
; PRIOR FILING DATE: 05-Sep-1997
; PRIOR APPLICATION NUMBER: 60/047,599
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,588
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,585
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,586
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,590
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,594
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,589
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,593
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,614
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/043,578
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/047,501
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/043,670
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/056,632
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,876
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,881
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,909
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,875
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,862
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,887
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,908
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/048,964
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/057,650
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/056,884
; PRIOR FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 280
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 64
; LENGTH: 2033

Query Match 76.0%; Score 15.2; DB 9; Length 2033;
Best Local Similarity 85.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ATATTAACGCCCACGCTCTC 20
||||||||| ||| | ||||
Db 43 ATATTAACGTCCATGTTCTC 24

RESULT 8
US-09-148-545-64/c
; Sequence 64, Application US/09148545
; Publication No. US20030027132A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 70 Human Secreted Proteins
; FILE REFERENCE: PI000193
; CURRENT APPLICATION NUMBER: US/09/148,545
; CURRENT FILING DATE: 1998-09-04
; EARLIER APPLICATION NUMBER: PCT/US98/04482
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,161
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23

; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,630
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,878
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,662
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,872
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,882
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,637
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,903
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,888
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,879
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,880
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,894
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,911
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,636
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,874
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,910
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,864
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,631
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,845
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,892
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/047,595
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/057,761
; EARLIER FILING DATE: 05-Sep-1997
; EARLIER APPLICATION NUMBER: 60/047,599
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,588
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,585
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,586
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,590
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,594
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,589
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,593
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,614
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,578
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,576
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/047,501
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,670
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/056,632
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,664
; EARLIER FILING DATE: 1997-08-22

```
;  EARLIER APPLICATION NUMBER: 60/056,876
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,881
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,909
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,875
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,862
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,887
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,908
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/048,964
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/057,650
;  EARLIER FILING DATE: 1997-09-05
;  EARLIER APPLICATION NUMBER: 60/056,884
;  EARLIER FILING DATE: 1997-08-22
;  NUMBER OF SEQ ID NOS: 280
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 64
;    LENGTH: 2033

  Query Match             76.0%;  Score 15.2;  DB 9;  Length 2033;
  Best Local Similarity   85.0%;  Pred. No. 1.3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCTC 20
           ||||||||| ||| | ||||
Db      43 ATATTAACGTCCATGTCTC 24


RESULT 9
US-09-981-876-120/c
; Sequence 120, Application US/09981876
; Patent No. US20020164669A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 70 Human Secreted Proteins
;  FILE REFERENCE: PI0001P1
;  CURRENT APPLICATION NUMBER: US/09/981,876
;  CURRENT FILING DATE:  2001-10-19
;  PRIOR APPLICATION NUMBER: 09/148,545
;  PRIOR FILING DATE: 1998-09-04
;  PRIOR APPLICATION NUMBER: 60/040,162
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: 60/040,333
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: 60/038,621
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: 60/040,161
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: 60/040,626
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: 60/040,334
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: 60/040,336
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: 60/040,163
;  PRIOR FILING DATE: 1997-03-07
;  PRIOR APPLICATION NUMBER: 60/047,615
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,600
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,597
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,502
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,633
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,583
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,617
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,618
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,503
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,592
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,581
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,584
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,500
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,587
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,492
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,598
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,613
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,582
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,596
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,612
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,632
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/047,601
;  PRIOR FILING DATE: 1997-05-23
;  PRIOR APPLICATION NUMBER: 60/043,580
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: 60/043,568
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: 60/043,314
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: 60/043,569
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: 60/043,311
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: 60/043,671
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: 60/043,674
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: 60/043,669
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: 60/043,312
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: 60/043,313
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: 60/043,672
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: 60/043,315
;  PRIOR FILING DATE: 1997-04-11
;  PRIOR APPLICATION NUMBER: 60/048,974
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/056,886
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: 60/056,877
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: 60/056,889
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: 60/056,893
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: 60/056,630
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: 60/056,878
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: 60/056,662
;  PRIOR FILING DATE: 1997-08-22
```

; PRIOR APPLICATION NUMBER: 60/056,872
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,882
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,637
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,903
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,888
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,879
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,880
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,894
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,911
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,636
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,874
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,910
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,864
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,631
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,845
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,892
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/047,595
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/057,761
; PRIOR FILING DATE: 05-Sep-1997
; PRIOR APPLICATION NUMBER: 60/047,599
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,588
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,585
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,586
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,590
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,594
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,589
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,593
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,614
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/043,578
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/047,501
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/043,670
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/056,632
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,876
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,881
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,909
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,875
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,862
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,887
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,908
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/048,964
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/057,650
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/056,884
; PRIOR FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 280
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 120
; LENGTH: 2136

Query Match 76.0%; Score 15.2; DB 9; Length 2136;
Best Local Similarity 85.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ATATTAACGCCCACGCTCTC 20
||||||||| ||| | ||||
Db 114 ATATTAACGTCCATGTTCTC 95

RESULT 10
US-09-148-545-120/c
; Sequence 120, Application US/09148545
; Publication No. US20030027132A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 70 Human Secreted Proteins
; FILE REFERENCE: PZ001P1
; CURRENT APPLICATION NUMBER: US/09/148,545
; CURRENT FILING DATE: 1998-09-04
; EARLIER APPLICATION NUMBER: PCT/US98/04482
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,161
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23

; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,630
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,878
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,662
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,872
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,882
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,637
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,903
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,888
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,879
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,880
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,894
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,911
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,636
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,874
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,910
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,864
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,631
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,845
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,892
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/047,595
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/057,761
; EARLIER FILING DATE: 05-Sep-1997
; EARLIER APPLICATION NUMBER: 60/047,599
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,588
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,585
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,586
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,590
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,594
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,589
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,593
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,614
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,578
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,576
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/047,501
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,670
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/056,632
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,664
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,876
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,881
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,909
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,875
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,862
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,887
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,908
; EARLIER FILING DATE: 1997-08-22

```
;  EARLIER APPLICATION NUMBER: 60/048,964
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/057,650
;  EARLIER FILING DATE: 1997-09-05
;  EARLIER APPLICATION NUMBER: 60/056,884
;  EARLIER FILING DATE: 1997-08-22
;  NUMBER OF SEQ ID NOS: 280
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 120
;    LENGTH: 2136

  Query Match             76.0%;  Score 15.2;  DB 9;  Length 2136;
  Best Local Similarity   85.0%;  Pred. No. 1.3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCTC 20
           ||||||||| ||| | ||||
Db     114 ATATTAACGTCCATGTTCTC 95


RESULT 11
US-09-820-893-44/c
; Sequence 44, Application US/09820893
; Patent No. US20020076705A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 31 Human Secreted Proteins
;  FILE REFERENCE: PZ033P1
;  CURRENT APPLICATION NUMBER: US/09/820,893
;  CURRENT FILING DATE:  2001-03-30
;  PRIOR APPLICATION NUMBER: 09/531,119
;  PRIOR FILING DATE: 2000-03-20
;  PRIOR APPLICATION NUMBER: 60/102,895
;  PRIOR FILING DATE: 1998-10-02
;  NUMBER OF SEQ ID NOS: 140
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
;    LENGTH: 2136
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-09-820-893-44

  Query Match             76.0%;  Score 15.2;  DB 10;  Length 2136;
  Best Local Similarity   85.0%;  Pred. No. 1.3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCTC 20
           ||||||||| ||| | ||||
Db     114 ATATTAACGTCCATGTTCTC 95


RESULT 12
US-09-917-800A-1469/c
; Sequence 1469, Application US/09917800A
; Patent No. US20020019462A1
; GENERAL INFORMATION:
;  APPLICANT: Mendrick, Donna
;  APPLICANT:  Porter, Mark
;  APPLICANT:  Johnson, Kory
;  APPLICANT:  Castle, Arthur
;  APPLICANT:  Elashoff, Michael
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Molecular Toxicology Modeling
;  FILE REFERENCE: 44921-5038-US
;  CURRENT APPLICATION NUMBER: US/09/917,800A
;  CURRENT FILING DATE:  2001-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,040
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,880
;  PRIOR FILING DATE: 2000-11-02
;  PRIOR APPLICATION NUMBER: US 60/290,029
;  PRIOR FILING DATE: 2001-05-11
;  PRIOR APPLICATION NUMBER: US 60/290,645
;  PRIOR FILING DATE: 2001-05-15
;  PRIOR APPLICATION NUMBER: US 60/292,336
;  PRIOR FILING DATE: 2001-05-22
;  PRIOR APPLICATION NUMBER: US 60/295,798
;  PRIOR FILING DATE: 2001-06-06
;  PRIOR APPLICATION NUMBER: US 60/297,457
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,884
;  PRIOR FILING DATE: 2001-06-19
;  PRIOR APPLICATION NUMBER: US 60/303,459
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 1740
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1469
;    LENGTH: 2196
;    TYPE: DNA
;    ORGANISM: Rattus norvegicus
;    FEATURE:
;    OTHER INFORMATION: Genbank Accession No. US20020119462A1 U63923
US-09-917-800A-1469

  Query Match             76.0%;  Score 15.2;  DB 10;  Length 2196;
  Best Local Similarity   85.0%;  Pred. No. 1.3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCTC 20
           || || ||||||||| ||||
Db    1054 ATCTTCACGCCCACGGTCTC 1035


RESULT 13
US-10-239-676-40/c
; Sequence 40, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
;  FILE REFERENCE: 5013.1003
;  CURRENT APPLICATION NUMBER: US/10/239,676
;  CURRENT FILING DATE:  2002-09-24
;  PRIOR APPLICATION NUMBER: PCT/EP01/03968
;     DE 10019058.8
;     DE 10019173.8
;     DE 10032529.7
;     DE 10043826.1
;  PRIOR FILING DATE: 2001-04-06
;     2000-04-05
;     2000-04-07
;     2000-06-30
;     2000-09-01
;  NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 40
;    LENGTH: 6061
;    TYPE: DNA
;    ORGANISM: Artificial Sequence
;    FEATURE:
;    OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-40

  Query Match             76.0%;  Score 15.2;  DB 9;  Length 6061;
  Best Local Similarity   85.0%;  Pred. No. 1.4e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCTC 20
           ||||||||| | ||| ||||
Db    2190 ATATTAACGTCAACGTTCTC 2171


RESULT 14
US-09-738-626-1562
; Sequence 1562, Application US/09738626
```

; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SHIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, MAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1562
; LENGTH: 873
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1562

Query Match 74.0%; Score 14.8; DB 9; Length 873;
Best Local Similarity 88.9%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TATTAACGCCCACGCTCT 19
|| |||||||||||| ||
Db 579 TAATACGCCCACGCGCT 596

RESULT 15
US-09-944-413-54/c
; Sequence 54, Application US/09944413
; Patent No. US20020156004A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Botstein,David
; APPLICANT: Eaton,Dan
; APPLICANT: Ferrara,Napoleone
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gerritsen,Mary
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul
; APPLICANT: Grimaldi,Christopher
; APPLICANT: Gurney,Austin
; APPLICANT: Hillan,Kenneth
; APPLICANT: Kljavin,Ivar
; APPLICANT: Napier,Mary
; APPLICANT: Roy,Margaret
; APPLICANT: Tumas,Daniel
; APPLICANT: Wood,William
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P2548P1C1
; CURRENT APPLICATION NUMBER: US/09/944,413
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 09/866,028
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: 60/067,411
; PRIOR FILING DATE: December 3, 1997
; PRIOR APPLICATION NUMBER: 60/069,334
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069335
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,278
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,425
; PRIOR FILING DATE: December 12, 1997
; PRIOR APPLICATION NUMBER: 60/069,696
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,694
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,702
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,870
; PRIOR FILING DATE: December 17, 1997
; PRIOR APPLICATION NUMBER: 60/069,873
; PRIOR FILING DATE: December 17, 1997
; PRIOR APPLICATION NUMBER: 60/068,017
; PRIOR FILING DATE: December 18, 1997
; PRIOR APPLICATION NUMBER: 60/070,440
; PRIOR FILING DATE: January 5, 1998
; PRIOR APPLICATION NUMBER: 60/074,086
; PRIOR FILING DATE: February 9, 1998
; PRIOR APPLICATION NUMBER: 60/074,092
; PRIOR FILING DATE: February 9, 1998
; PRIOR APPLICATION NUMBER: 60/075,945
; PRIOR FILING DATE: February 25, 1998
; PRIOR APPLICATION NUMBER: 60/112,850
; PRIOR FILING DATE: December 16, 1998
; PRIOR APPLICATION NUMBER: 60/113,296
; PRIOR FILING DATE: December 22, 1998
; PRIOR APPLICATION NUMBER: 60/146,222
; PRIOR FILING DATE: July 28, 1999
; PRIOR APPLICATION NUMBER: PCT/US98/19330
; PRIOR FILING DATE: September 16, 1998
; PRIOR APPLICATION NUMBER: PCT/US98/25108
; PRIOR FILING DATE: December 1, 1998
; PRIOR APPLICATION NUMBER: 09/216,021
; PRIOR FILING DATE: December 16, 1998
; PRIOR APPLICATION NUMBER: 09/218,517
; PRIOR FILING DATE: December 22, 1998
; PRIOR APPLICATION NUMBER: 09/254,311
; PRIOR FILING DATE: March 3, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: June 22, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: September 15, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28409
; PRIOR FILING DATE: No. US20020156004Alember 30, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: No. US20020156004Alember 30, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28301
; PRIOR FILING DATE: Decemberl, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: December 16, 1999
; PRIOR APPLICATION NUMBER: PCT/US00/03565
; PRIOR FILING DATE: February 11, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: February 22, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: March 2, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: March 30, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/14042
; PRIOR FILING DATE: May 22, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/20710
; PRIOR FILING DATE: July 28, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: December 1, 2000
; PRIOR APPLICATION NUMBER: PCT/US01/06520
; PRIOR FILING DATE: February 28, 2001
; NUMBER OF SEQ ID NOS: 120
; SEQ ID NO 54
; LENGTH: 2331
; TYPE: DNA
; ORGANISM: Homo Sapien

US-09-944-413-54

Query Match 74.0%; Score 14.8; DB 9; Length 2331;
Best Local Similarity 88.9%; Pred. No. 2.1e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATATTAACGCCCACGCTC 18
||||||:||| ||||||||
Db 643 ATATTCACGTCCACGCTC 626

Search completed: May 12, 2003, 04:49:52
Job time : 20.7143 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 01:04:28 ; Search time 244.143 Seconds
(without alignments)
1326.722 Million cell updates/sec

Title: US-09-980-277-3
Perfect score: 20
Sequence: 1 atattaacgcccacgctctc 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	20	100.0	687	17	BH210086	BH210086 Sm1-56J20
2	17.4	87.0	570	17	AQ464286	AQ464286 HS_5101_A
c 3	16.8	84.0	319	17	AZ216980	AZ216980 Sheared D
c 4	16.8	84.0	706	12	BG429009	BG429009 602501922
c 5	16.8	84.0	804	10	AV755514	AV755514 AV755514
c 6	16.4	82.0	219	9	AA698952	AA698952 HL06045.5
c 7	16.4	82.0	294	17	BH809948	BH809948 SALK_0368
c 8	16.4	82.0	307	9	AA698954	AA698954 HL06047.5
c 9	16.4	82.0	573	13	BI962036	BI962036 01231 lea
c 10	16.4	82.0	679	13	BM133610	BM133610 NXLV_009_
c 11	16.4	82.0	700	9	AA940956	AA940956 LD23315.5
c 12	16.4	82.0	872	9	AA950399	AA950399 LD28939.5
c 13	16.4	82.0	1503	17	AQ752099	AQ752099 HS_5570_B
c 14	16	80.0	940	17	CN801MAP	AL150690 Anopheles
15	16	80.0	1161	14	BQ067992	BQ067992 AGENCOURT
16	15.8	79.0	382	12	BG639617	BG639617 SD03435.3
c 17	15.8	79.0	400	9	AI210561	AI210561 j9b05a1.r
18	15.8	79.0	401	14	Z14884	Z14884 CEL7H1 Chri
c 19	15.8	79.0	441	17	AQ908431	AQ908431 Gssftc0518
20	15.8	79.0	459	17	AQ707815	AQ707815 HS_5561_B
21	15.8	79.0	487	17	AQ406537	AQ406537 HS_2262_B
22	15.8	79.0	491	17	AZ183441	AZ183441 SP_0192_A
23	15.8	79.0	514	9	AU208418	AU208418 AU208418
24	15.8	79.0	605	13	BJ119342	BJ119342 BJ119342
25	15.8	79.0	621	17	AZ226253	AZ226253 RPCI-23-9
26	15.8	79.0	644	12	BG851049	BG851049 1024030B0
c 27	15.8	79.0	664	17	AZ870610	AZ870610 2M0183H12
28	15.8	79.0	755	13	BI524144	BI524144 603051554
c 29	15.8	79.0	838	9	AI648350	AI648350 AEMTAE78
c 30	15.8	79.0	897	12	BG287095	BG287095 602382396
c 31	15.8	79.0	931	12	BF141420	BF141420 601786358
32	15.8	79.0	947	17	AQ099094	AQ099094 Pan trogl
33	15.8	79.0	973	17	CNS02QAT	AL209054 Tetraodon
34	15.8	79.0	1014	17	AG111100	AG111100 Pan trogl
35	15.8	79.0	1028	17	AG097971	AG097971 Pan trogl
36	15.8	79.0	1112	12	BF686423	BF686423 602143210
c 37	15.8	79.0	1179	14	BQ953557	BQ953557 AGENCOURT
38	15.8	79.0	1650	14	BQ706996	BQ706996 AGENCOURT
c 39	15.4	77.0	235	17	AZ910305	AZ910305 RPCI-24-8
c 40	15.4	77.0	320	12	BF339308	BF339308 602038636
c 41	15.4	77.0	394	17	AQ084176	AQ084176 HS_2244_A
c 42	15.4	77.0	421	10	AV789035	AV789035 AV789035
43	15.4	77.0	441	12	BF042015	BF042015 BP250023A
44	15.4	77.0	441	12	BF044104	BF044104 BP250023A
45	15.4	77.0	667	17	AG101516	AG101516 Pan trogl

ALIGNMENTS

RESULT 1
BH210086/c
LOCUS BH210086 687 bp DNA linear GSS 24-OCT-2001
DEFINITION Sm1-56J20.TF Sm1 Schistosoma mansoni genomic clone Sm1-56J20, DNA sequence.
ACCESSION BH210086
VERSION BH210086.1 GI:16389299
KEYWORDS GSS.
SOURCE Schistosoma mansoni.
ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE 1 (bases 1 to 687)
AUTHORS Shetty,J., Simpson,A., Malek,J., Koo,H., LoVerde,P.T. and El-Sayed,N.M.
TITLE Use of end sequences from Schistosoma mansoni (Puerto Rico strain) Sm1 BAC library for gene discovery and map construction
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: Sm1-56J20.TR
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
lo.edu).
Seq primer: M13 For
Class: BAC ends.

```
FEATURES             Location/Qualifiers
     source          1. .687
                     /organism="Schistosoma mansoni"
                     /strain="Puerto Rico"
                     /db_xref="taxon:6183"
                     /clone="Sm1-56J20"
                     /clone_lib="Sm1"
                     /note="Vector: pBeloBAC11; Site_1: Hin dIII; Constructed
                     in the laboratory of Dr. Denis Le Paslier at the Fondation
                     Jean Dausset, CEPH, Paris, France. Briefly, Schistosoma
                     mansoni agarose embedded DNA was partially digested with
                     Hin dIII. High molecular weight fragments were ligated in
                     pBeloBAC11 digested with Hin dIII. The average insert size
                     is 100 Kb. Total clone coverage: approx. 7.95 X the
                     haploid genome. Further information can be found in Le
                     Paslier et al. (2000) Construction and characterization of
                     a Schistosoma mansoni bacterial artificial chromosome
                     library. Genomics 65: 87-94."
BASE COUNT       218 a    101 c    152 g    216 t
ORIGIN

  Query Match          100.0%;  Score 20;  DB 17;  Length 687;
  Best Local Similarity  100.0%;  Pred. No. 20;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTCTC 20
           ||||||||||||||||||||
Db     336 ATATTAACGCCCACGCTCTC 317


RESULT 2
AQ464286
LOCUS       AQ464286               570 bp    DNA     linear   GSS 23-APR-1999
DEFINITION  HS_5101_A1_B08_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=677 Col=15 Row=C, DNA sequence.
ACCESSION   AQ464286
VERSION     AQ464286.1  GI:4641381
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 677  row: C  column: 15
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 570.
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=677 Col=15 Row=C"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI:
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT       167 a    204 c     63 g    127 t      9 others
ORIGIN

  Query Match          87.0%;  Score 17.4;  DB 17;  Length 570;
  Best Local Similarity  94.7%;  Pred. No. 3.5e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 TATTAACGCCCACGCTCTC 20
           ||||||| ||||||||||||
Db     212 TATTAACGTCCACGCTCTC 230


RESULT 3
AZ216980/c
LOCUS       AZ216980               319 bp    DNA     linear   GSS 09-JUN-2000
DEFINITION  Sheared DNA-57G10.TR Sheared DNA Trypanosoma brucei genomic clone
            Sheared DNA-57G10, DNA sequence.
ACCESSION   AZ216980
VERSION     AZ216980.1  GI:8434780
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 319)
  AUTHORS   El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
            Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.
            , Fraser,C. and Adams,M.
  TITLE     Determination of clone end sequences from Trypanosoma brucei GUTat
            10.1 sheared DNA library
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
            DNA library constructed at TIGR. Clones will be available for
            distribution through Research Genetics, Alabama, DSA. Sheared DNA
            end sequences search page: http://www.tigr.org/tdb/mdb/tbdb/.
            Seq primer: M13-Reverse
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .319
                     /organism="Trypanosoma brucei"
                     /strain="TREU927/4 GUTat 10.1"
                     /db_xref="taxon:5691"
                     /clone="Sheared DNA-57G10"
                     /clone_lib="Sheared DNA"
                     /note="Vector: pUC18; Site_1: SmaI; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from a cloned population of
                     Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
                     sheared to give a tight size distribution (approx 2 kb).
                     The v + i method used for the library construction is
                     described in detail in Smith, H.O. and Venter, J.C.
                     (Making small insert libraries for whole genome shotgun
                     sequencing projects.  In Genome Sequencing: A Practical
                     Approach, eds. M. Vaudin and B. Barell, Oxford University
                     Press, 1999)."
BASE COUNT        54 a     56 c     93 g    116 t
ORIGIN

  Query Match          84.0%;  Score 16.8;  DB 17;  Length 319;
  Best Local Similarity  90.0%;  Pred. No. 5.8e+02;
```

```
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ATATTAACGCCCACGCTCTC 20
          |||||||||| ||| |||||
Db    188 ATATTAACGCACACACTCTC 169


RESULT 4
BG429009/c
LOCUS       BG429009                706 bp    mRNA    linear   EST 14-MAR-2001
DEFINITION  602501922F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4615444 5',
            mRNA sequence.
ACCESSION   BG429009
VERSION     BG429009.1  GI:13335515
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 706)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1368  row: 1  column: 05
            High quality sequence stop: 676.
FEATURES             Location/Qualifiers
     source          1..706
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4615444"
                     /clone_lib="NIH_MGC_75"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
                     sfiI (ggccgcctcggcc); Site_2: sfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T).  Average insert size 1.65
                     kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      174 a    158 c    191 g    182 t      1 others
ORIGIN

  Query Match           84.0%;  Score 16.8;  DB 12;  Length 706;
  Best Local Similarity   90.0%;  Pred. No. 7.2e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ATATTAACGCCCACGCTCTC 20
          |||||||||||| | ||||
Db     41 ATATTAACGCCCATGCTCTC 22


RESULT 5
AV755514/c
LOCUS       AV755514                804 bp    mRNA    linear   EST 19-OCT-2000
DEFINITION  AV755514 BM Homo sapiens cDNA clone BMPBGD11 5', mRNA sequence.
ACCESSION   AV755514
VERSION     AV755514.1  GI:10913362
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 804)
  AUTHORS   Gu,J., Zhao,X., Huang,Q., Xu,X., Li,Y., Peng,Y., Song,H., Xiao,H.,
            Gu,Y., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng
            ,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G.,
            Yang,Y., Gao,G., Wang,Z., Zhang,Q., Chen,S., Han,Z. and Chen,Z.
  TITLE     Homo sapiens cDNA BM clones
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801949(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1..804
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="BMPBGD11"
                     /clone_lib="BM"
                     /tissue_type="Bone marrow"
                     /cell_type="CD34+ hematopoietic stem/progenitor cell"
                     /lab_host="BM25.8"
                     /note="Vector: pTriplEx2; Site_1: sfiIA; Site_2: sfiIB"
BASE COUNT      177 a    189 c    221 g    212 t      5 others
ORIGIN

  Query Match           84.0%;  Score 16.8;  DB 10;  Length 804;
  Best Local Similarity   90.0%;  Pred. No. 7.5e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ATATTAACGCCCACGCTCTC 20
          |||||||| ||||| ||||
Db    119 ATATTAACGTCCACGTCTC 100


RESULT 6
AA698952/c
LOCUS       AA698952                219 bp    mRNA    linear   EST 19-APR-2001
DEFINITION  HL06045.5prime HL Drosophila melanogaster head BlueScript
            Drosophila melanogaster cDNA clone HL06045 5prime, mRNA sequence.
ACCESSION   AA698952
VERSION     AA698952.1  GI:2701881
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 219)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 60  row: D  column: 9
            High quality sequence stop: 157.
FEATURES             Location/Qualifiers
     source          1..219
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="HL06045"
                     /clone_lib="HL Drosophila melanogaster head BlueScript"
                     /sex="male and female"
                     /dev_stage="adult"
                     /lab_host="SOLR"
```

```
                     /note="Organ: head--brain & sensory organ; Vector:
                     BlueScript SK; Site_1: EcoRI; Site_2: XhoI; Constructed
                     using Stratagene ZAP-cDNA Synthesis kit. Oligo dT-primed
                     and directionally cloned at EcoRI and XhoI in BlueScript
                     SK(+/-)"
BASE COUNT        48 a     54 c     58 g     59 t
ORIGIN

  Query Match           82.0%;  Score 16.4;  DB 9;  Length 219;
  Best Local Similarity   94.4%;  Pred. No. 8.3e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 ATTAACGCCCACGCTCTC 20
           |||||||||| |||||||
Db      96 ATTAACGCCCCCGCTCTC 79


RESULT 7
BH809948/c
LOCUS       BH809948                294 bp    DNA     linear   GSS 02-MAY-2002
DEFINITION  SALK_036819 Arabidopsis thaliana TDNA insertion lines Arabidopsis
            thaliana genomic clone SALK_036819, DNA sequence.
ACCESSION   BH809948
VERSION     BH809948.1  GI:20387765
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 294)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
            ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
            , Zimmerman,J. and Ecker,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .294
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="SALK_036819"
                     /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                     /note="PCR was performed on Arabidopsis thaliana lines
                     each of which contains  one or more TDNA insertion
                     elements. The resultant fragment for each  line was
                     directly sequenced to determine the genomic sequence at
                     the  site of insertion. Details of the protocols used can
                     be found at  http://signal.salk.edu/tdna_protocols.html"
BASE COUNT        81 a     69 c     53 g     91 t
ORIGIN

  Query Match           82.0%;  Score 16.4;  DB 17;  Length 294;
  Best Local Similarity   94.4%;  Pred. No. 9e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCACGCTC 18
           ||||||||||| ||||||
Db     205 ATATTAACGCCAACGCTC 188


RESULT 8
```

```
AA698954/c
LOCUS       AA698954                307 bp    mRNA    linear   EST 19-APR-2001
DEFINITION  HL06047.5prime HL Drosophila melanogaster head BlueScript
            Drosophila melanogaster cDNA clone HL06047 5prime, mRNA sequence.
ACCESSION   AA698954
VERSION     AA698954.1  GI:2701883
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 307)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 60  row: D  column: 11
            High quality sequence stop: 219.
FEATURES             Location/Qualifiers
     source          1. .307
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="HL06047"
                     /clone_lib="HL Drosophila melanogaster head BlueScript"
                     /sex="male and female"
                     /dev_stage="adult"
                     /lab_host="SOLR"
                     /note="Organ: head--brain & sensory organ; Vector:
                     BlueScript SK; Site_1: EcoRI; Site_2: XhoI; Constructed
                     using Stratagene ZAP-cDNA Synthesis kit. Oligo dT-primed
                     and directionally cloned at EcoRI and XhoI in BlueScript
                     SK(+/-)"
BASE COUNT        73 a     72 c     85 g     77 t
ORIGIN

  Query Match           82.0%;  Score 16.4;  DB 9;  Length 307;
  Best Local Similarity   94.4%;  Pred. No. 9.1e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 ATTAACGCCCACGCTCTC 20
           |||||||||| |||||||
Db      96 ATTAACGCCCCCGCTCTC 79


RESULT 9
BI962036/c
LOCUS       BI962036                573 bp    mRNA    linear   EST 22-OCT-2001
DEFINITION  01231 leafy spurge Lambda HybriZAP 2.1 two-hybrid vector cDNA
            Library Euphorbia esula cDNA clone 16AR 5', mRNA sequence.
ACCESSION   BI962036
VERSION     BI962036.1  GI:16320239
KEYWORDS    EST.
SOURCE      leafy spurge.
  ORGANISM  Euphorbia esula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Malpighiales; Euphorbiaceae; Euphorbia.
REFERENCE   1  (bases 1 to 573)
  AUTHORS   Anderson,J.V. and Horvath,D.P.
  TITLE     Identification of mRNAs expressed in underground adventitious buds
            of Euphorbia esula (leafy spurge)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Anderson JV
            Plants Science Research
            USDA/ARS, Biosciences Research Lab
            1605 Albrecht Blvd., PO Box 5674, Fargo, ND 58105, USA
```

```
            Tel: 701 239 1263
            Fax: 701 239 1252
            Email: andersjv@fargo.ars.usda.gov
            Seq primer: pAD5.
FEATURES              Location/Qualifiers
     source           1. .573
                      /organism="Euphorbia esula"
                      /db_xref="taxon:3993"
                      /clone="16AR"
                      /clone_lib="leafy spurge Lambda HybriZAP 2.1 two-hybrid
                      vector cDNA Library"
                      /tissue_type="underground adventitious buds"
                      /dev_stage="3-day induced (decapitated)"
BASE COUNT      134 a    152 c    129 g    155 t      3 others
ORIGIN

  Query Match           82.0%;  Score 16.4;  DB 13;  Length 573;
  Best Local Similarity   85.0%;  Pred. No. 1.1e+03;
  Matches   17;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ATATTAACGCCCACGCTCTC 20
          |||||||||||| ||||||||
Db    299 AGATTAACGCCCAACCTCTC 280


RESULT 10
BM133610/c
LOCUS       BM133610                679 bp    mRNA    linear   EST 07-JAN-2002
DEFINITION  NXLV_009_F09_F NXLV (Nsf Xylem Late wood Vertical) Pinus taeda cDNA
            clone NXLV_009_F09 5', mRNA sequence.
ACCESSION   BM133610
VERSION     BM133610.1  GI:17140689
KEYWORDS    EST.
SOURCE      loblolly pine.
  ORGANISM  Pinus taeda
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE   1  (bases 1 to 679)
  AUTHORS   Sederoff,R.
  TITLE     Molecular Basis of Wood Formation in the Pine Megagenome
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Johnson, Arthur
            North Carolina State University
            Tel: 919 515 7800
            Fax: 919 515 7801
            Email: ajohnson@unity.ncsu.edu
            Seq primer: T3.
FEATURES              Location/Qualifiers
     source           1. .679
                      /organism="Pinus taeda"
                      /strain="Coastal plain loblolly pine from North Carolina"
                      /db_xref="taxon:3352"
                      /clone="NXLV_009_F09"
                      /clone_lib="NXLV (Nsf Xylem Late wood Vertical)"
                      /tissue_type="primary xylem"
                      /dev_stage="late wood"
                      /lab_host="XL1-Blue"
                      /note="Vector: pTriplEx; Site_1: EcoRI; The library is
                      from late (summer-August) wood, taken from below the crown
                      of a 20 year old tree. The harvested xylem tissue was on
                      the the cusp between transitional and mature wood. NOTE:
                      The sequences contain a 'cDNA adapter' between the EcoRI
                      site and the start of the EST. The adapter sequence is
                      'AATTCGGCCATTAIGGCC'."
BASE COUNT      177 a    101 c    174 g    210 t      17 others
ORIGIN

  Query Match           82.0%;  Score 16.4;  DB 13;  Length 679;
  Best Local Similarity   89.5%;  Pred. No. 1.1e+03;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 TATTAACGCCCACGCTCTC 20
          ||||||||||| |||||||
Db    558 TATTAACGCCAACGCTCTC 540


RESULT 11
AA940956/c
LOCUS       AA940956                700 bp    mRNA    linear   EST 23-APR-2001
DEFINITION  LD23315.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD23315 5 similar to CG8240: FBan0008240
            'signal transduction' located on: X 16C1-16C1;: 04/10/2001, mRNA
            sequence.
ACCESSION   AA940956
VERSION     AA940956.2  GI:13767616
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 700)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     On May 1, 1998 this sequence version replaced gi:3100869.
            Other_ESTs: LD23315.3prime
            Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Tel: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003506: arm:X [17291364,17592357]
            estimated-cyto:16B4-16D6: 04/10/2001
            Plate: LD.233  row: B  column: 3
            High quality sequence stop: 595
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .700
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone="LD23315"
                      /clone_lib="LD Drosophila melanogaster embryo pOT2"
                      /sex="male and female"
                      /dev_stage="0 to 24 hours mixed stage embryonic"
                      /lab_host="XL1 Blue"
                      /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                      XhoI; Sized fractionated cDNAs were directly ligated into
                      pOT2. "
BASE COUNT      171 a    178 c    175 g    176 t
ORIGIN

  Query Match           82.0%;  Score 16.4;  DB 9;  Length 700;
  Best Local Similarity   94.4%;  Pred. No. 1.1e+03;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 ATTAACGCCCACGCTCTC 20
          |||||||||| |||||||
Db    589 ATTAACGCCCCCGCTCTC 572


RESULT 12
AA950399/c
LOCUS       AA950399                872 bp    mRNA    linear   EST 23-APR-2001
DEFINITION  LD28939.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD28939 5 similar to CG8240: FBan0008240
            'signal transduction' located on: X 16C1-16C1;: 04/10/2001, mRNA
            sequence.
ACCESSION   AA950399
VERSION     AA950399.2  GI:13768163
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
```

```
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 972)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     On May 4, 1998 this sequence version replaced gi:3111771.
            Other_ESTs: LD28939.3prime
            Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003506: arm:X [17291364,17592357]
            estimated-cyto:16B4-16D6; 04/10/2001
            Plate: LD.289  row: D  column: 3
            High quality sequence stop: 739
            POLYA-No.
FEATURES             Location/Qualifiers
     source          1. .872
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="LD28939"
                     /clone_lib="LD Drosophila melanogaster embryo pOT2"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="XL1 Blue"
                     /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2. "
BASE COUNT       220 a    224 c    224 g    204 t
ORIGIN

  Query Match          82.0%;  Score 16.4;  DB 9;  Length 872;
  Best Local Similarity   94.4%;  Pred. No. 1.2e-03;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 ATTAACGCCCACGCTCTC 20
           ||||||||||| ||||||
Db     592 ATTAACGCCCCGCTCTC 575


RESULT 13
AQ752099/c
LOCUS       AQ752099                1503 bp    DNA     linear   GSS 19-JUL-1999
DEFINITION  HS_5570_B2_G10_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1146 Col=20 Row=N, DNA sequence.
ACCESSION   AQ752099
VERSION     AQ752099.1  GI:5539257
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1503)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Purlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 1146  row: N  column: 20
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 1503.
FEATURES             Location/Qualifiers
     source          1. .1503
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate-1146 Col-20 Row-N"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT       452 a    311 c    520 g    197 t     23 others
ORIGIN

  Query Match          82.0%;  Score 16.4;  DB 17;  Length 1503;
  Best Local Similarity   94.4%;  Pred. No. 1.4e+03;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 ATTAACGCCCACGCTCTC 20
           ||||||||||| |||||
Db     206 ATTAACGCCCACCCTCTC 189


RESULT 14
CNS01MAP/c
LOCUS       CNS01MAP                940 bp    DNA     linear   GSS 14-JUN-2001
DEFINITION  Anopheles gambiae GSS T7 end of clone 21P09 of NotreDamel library
            from strain PEST of Anopheles gambiae (African malaria mosquito),
            genomic survey sequence.
ACCESSION   AL150690
VERSION     AL150690.1  GI:7011169
KEYWORDS    GSS.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 940)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
REFERENCE   2  (bases 1 to 940)
  AUTHORS   Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
            Roux, Paris 75015, France
COMMENT     This clone is from an A. gambiae BAC library provided by F.H.
            Collins and sequenced by Genoscope in collaboration with the
            Laboratory of Biochem. and Biol. Molec. of Insects, Institut
            Pasteur.
FEATURES             Location/Qualifiers
     source          1. .940
                     /organism="Anopheles gambiae"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="21P09"
                     /clone_lib="NotreDamel"
                     /note="end : T7"
BASE COUNT       219 a    131 c    276 g    287 t     27 others
ORIGIN

  Query Match          80.0%;  Score 16;  DB 17;  Length 940;
```

```
  Best Local Similarity   80.0%;  Pred. No. 1.9e+03;
  Matches   16;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ATATTAACGCCCAGCTCTC 20
           || :|||||||||| |||||:
Db     570 ATTYTAACGCCCACTCTCTT 551


RESULT 15
BQ067992
LOCUS       BQ067992               1161 bp    mRNA    linear   EST 02-APR-2002
DEFINITION  AGENCOURT_6792474 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5768997
            5', mRNA sequence.
ACCESSION   BQ067992
VERSION     BQ067992.1  GI:19897038
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1161)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12829  row: l  column: 22
            High quality sequence stop: 440.
FEATURES             Location/Qualifiers
     source          1..1161
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5768997"
                     /clone_lib="NIH_MGC_121"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 3
                     fetal brains, female age 20 weeks, female age 24 weeks,
                     and male age 26 weeks.  Library is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed  upon
                     cloning).  Average insert size 1.7 kb, insert size range
                     0.7-3.5 kb.  Library is normalized and enriched for
                     full-length clones and  was constructed by C. Gruber
                     (Invitrogen).  Research Genetics  tracking code 017. Note:
                     this is a NIH_MGC Library."
BASE COUNT      158 a    474 c    334 g    195 t
ORIGIN

  Query Match             80.0%;  Score 16;  DB 14;  Length 1161;
  Best Local Similarity   100.0%;  Pred. No. 2e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 TAACGCCCAGCTCTC 20
           ||||||||||||||||
Db     919 TAACGCCCACGCTCTC 934


Search completed: May 12, 2003, 02:35:28
Job time : 249.143 secs
```